SCIENTIFIC AMERICAN Resource Library

Physical Sciences 1

SCIENTIFIC AMERICAN Resource Library

READINGS IN THE

Physical Sciences VOLUME 1

OFFPRINTS 201–246

with an Introduction by Isaac Asimov
Boston University School of Medicine

W. H. FREEMAN AND COMPANY San Francisco

Each of the articles in this volume is available as a separate Offprint. For a complete listing of Offprints available in the Life Sciences, Chemistry, Physics, Technology, Psychology, the Social Sciences, the History and Philosophy of Science, and the Earth Sciences, write to W. H. Freeman and Company, 660 Market Street, San Francisco, California 94104 or to W. H. Freeman and Company, Ltd., Warner House, Folkstone, Kent, England.

Printed in the United States of America.
Library of Congress Catalog Card Number: 78-87738
Standard Book Number: 7167 0991-0

Organization of the SCIENTIFIC AMERICAN Resource Library

Subject Series:

The *SCIENTIFIC AMERICAN Resource Library* is a multi-volume compilation of more than 700 articles selected from the magazine. These are organized under five subject classifications.

Readings in Earth Sciences (2 volumes)
Readings in Life Sciences (7 volumes)
Readings in Physical Sciences and Technology (3 volumes)
Readings in Psychology (2 volumes)
Readings in Social Sciences (1 volume)

Numbering System:

Each article is numbered and the articles in the volumes are arranged in numerical order. The numbers assigned to each series are:

Earth Sciences, 801-999.
Life Sciences, 1-199 and 1001-1999.
Physical Sciences and Technology, 201-399.
Psychology, 401-599.
Social Sciences, 601-799.

Topic Index:

This index classifies the Readings in Physical Sciences by topic. Note that articles from other subject series that are relevant to the topic are also listed.

Author Index:

The authors of the articles in all five subject series are given. The numbers after the authors' names are the article numbers, not page numbers.

Scientific American Offprints:

Every article in these volumes is published separately in the SCIENTIFIC AMERICAN Offprint Series and may be purchased in any quantity. Order by article number and title.
The number of each offprint corresponds to the number on each article in the Resource Library.
A catalog of SCIENTIFIC AMERICAN offprints may be obtained from the publisher.

Additions to the Series:

New titles are added to the SCIENTIFIC AMERICAN Offprint Series every month and when enough new articles on a subject are available, a new bound volume will be added to the *SCIENTIFIC AMERICAN Resource Library.*

W. H. Freeman and Company
660 Market Street, San Francisco, California 94104
Warner House, Folkestone, Kent, England

Contents

VOLUME **1**

VOLUME 2

VOLUME 3

Topic Index

(This index includes not only the article in the Physical Sciences but also relevant articles from other subject series.)

MATTER IN BULK

NUCLEUS AND ATOM

ORGANIC AND BIOCHEMISTRY

1064. Franklin C. McLean BONE
1068. N. W. Pirie ORTHODOX AND UNORTHODOX METHODS OF MEETING WORLD FOOD NEEDS
1074. Charles Yanofsky GENE STRUCTURE AND PROTEIN STRUCTURE
1075. Ruth Hubbard & Allen Kropf MOLECULAR ISOMERS IN VISION
1076. Paul R. Ehrlich & Peter H. Raven BUTTERFLIES AND PLANTS
1077. Bernard W. Agranoff MEMORY AND PROTEIN SYNTHESIS
1078. Carroll M. Williams THIRD-GENERATION PESTICIDES
1081. Elijah Adams BARBITURATES
1082. Trevor Robinson ALKALOIDS
1083. R. R. Porter THE STRUCTURE OF ANTIBODIES
1085. Anthony Allison LYSOSOMES AND DISEASE
1087. Tsutomu Watanabe INFECTIOUS DRUG RESISTANCE
1092. Brian F. C. Clark & Kjeld A. Marcker HOW PROTEINS START
1101. Efraim Racker THE MEMBRANE OF THE MITOCHONDRION
1103. Leonard Hayflick HUMAN CELLS AND AGING
1110. R. D. Preston PLANTS WITHOUT CELLULOSE
1111. Johannes van Overbeek THE CONTROL OF PLANT GROWTH
1122. Sir Solly Zuckerman HORMONES
1124. Arthur Kornberg THE SYNTHESIS OF DNA
280. John F. Brown, Jr. INCLUSION COMPOUNDS
286. Edel Wasserman CHEMICAL TOPOLOGY
297. John E. Amoore, James W. Johnston, Jr., & Martin Rubin THE STEREOCHEMICAL THEORY OF ODOR
304. Marshall Gates ANALGESIC DRUGS
308. Geoffrey Eglinton & Melvin Calvin CHEMICAL FOSSILS
314. Herman F. Mark GIANT MOLECULES
315. Giulio Natta PRECISELY CONSTRUCTED POLYMERS
320. R. B. Merrifield THE AUTOMATIC SYNTHESIS OF PROTEINS
325. Sir Lawrence Bragg X-RAY CRYSTALLOGRAPHY
446. Harold E. Himwich THE NEW PSYCHIATRIC DRUGS
483. Frank Barron, Murray E. Jarvik & Sterling Bunnell, Jr. THE HALLUCINOGENIC DRUGS
497. Lawrence Zelic Freedman "TRUTH" DRUGS

PHYSICAL CHEMISTRY: LIQUIDS AND SOLIDS

249. Gregory H. Wannier THE NATURE OF SOLIDS
258. George A. W. Boehm TITANIUM
259. Stephen M. Shelton ZIRCONIUM
260. Robert L. Fullman THE GROWTH OF CRYSTALS
262. Arthur M. Buswell & Worth H. Rodebush WATER
263. Charles M. Herzfeld & Arnold M. Bass FROZEN FREE RADICALS
265. B. J. Adler & Thomas E. Wainwright MOLECULAR MOTIONS
266. Boris V. Derjaguin THE FORCE BETWEEN MOLECULES
267. J. D. Bernal THE STRUCTURE OF LIQUIDS
288. Robert L. Sproull THE CONDUCTION OF HEAT IN SOLIDS
307. L. K. Runnels ICE
854. Edwin Roedder ANCIENT FLUIDS IN CRYSTALS

QUANTUM THEORY AND RELATIVITY

205. Karl K. Darrow THE QUANTUM THEORY
208. Freeman J. Dyson FIELD THEORY
209. Albert Einstein ON THE GENERALIZED THEORY OF GRAVITATION
212. George Gamow THE PRINCIPLE OF UNCERTAINTY
231. Philip Morrison THE OVERTHROW OF PARITY
241. Erwin Schrödinger WHAT IS MATTER?
264. George Gamow THE EXCLUSION PRINCIPLE
273. George Gamow GRAVITY
291. J. Bronowski THE CLOCK PARADOX
292. P. A. M. Dirac THE EVOLUTION OF THE PHYSICIST'S PICTURE OF NATURE
301. Eugene P. Wigner VIOLATIONS OF SYMMETRY IN PHYSICS
309. Martin Gardner CAN TIME GO BACKWARD?
316. Kip S. Thorne GRAVITATIONAL COLLAPSE
321. R. S. Shankland THE MICHELSON-MORLEY EXPERIMENT

TECHNOLOGY

187. Derek H. Fender CONTROL MECHANISMS OF THE EYE
190. Maynard A. Amerine WINE
1020. Alfred Champagnat PROTEIN FROM PETROLEUM
1023. Willem J. Kolff AN ARTIFICIAL HEART INSIDE THE BODY
1036. Joseph B. MacInnis LIVING UNDER THE SEA
1040. Robert S. Ledley & Frank H. Ruddle CHROMOSOME ANALYSIS BY COMPUTER
1043. Cyrus Levinthal MOLECULAR MODEL-BUILDING BY COMPUTER
215. James P. Gordon THE MASER
221. Henry F. Ivey ELECTROLUMINESCENCE
222. Abram F. Joffe THE REVIVAL OF THERMOELECTRICITY
225. Harold Lyons ATOMIC CLOCKS
236. Richard F. Post FUSION POWER
237. J. S. Prener & D. B. Sullenger PHOSPHORS
246. Lyman Spitzer, Jr. THE STELLARATOR
270. Theodore A. Buchhold APPLICATIONS OF SUPERCONDUCTIVITY
274. Arthur L. Schawlow OPTICAL MASERS
279. J. E. Kunzler & Morris Tanenbaum SUPERCONDUCTING MAGNETS
281. H. S. Feder & A. E. Spencer TELEPHONE SWITCHING
284. Olexa-Myron Bilaniuk SEMICONDUCTOR PARTICLE-DETECTORS
288. Robert L. Sproull THE CONDUCTION OF HEAT IN SOLIDS
294. Arthur L. Schawlow ADVANCES IN OPTICAL MASERS
298. William H. Pierce REDUNDANCY IN COMPUTERS
299. Gerald Oster & Yasunori Nishijima MOIRE PATTERNS
300. Emmett N. Leith & Juris Upatnieks PHOTOGRAPHY BY LASER
302. Stewart E. Miller COMMUNICATION BY LASER
303. George C. Pimentel CHEMICAL LASERS
306. Leo L. Beranek NOISE
310. Werner H. Wahl & Henry H. Kramer NEUTRON-ACTIVATION ANALYSIS
312. Anatol Rapoport THE USE AND MISUSE OF GAME THEORY
313. Marvin L. Minsky ARTIFICIAL INTELLIGENCE
314. Herman F. Mark GIANT MOLECULES
315. Giulio Natta PRECISELY CONSTRUCTED POLYMERS
318. Bela Julesz TEXTURE AND VISUAL PERCEPTION
319. Jerome B. Wiesner & Herbert F. York NATIONAL SECURITY AND THE NUCLEAR-TEST BAN
326. Allen V. Astin STANDARDS OF MEASUREMENT
503. Burke M. Smith THE POLYGRAPH
510. Oliver G. Selfridge & Ulric Neisser PATTERN RECOGNITION BY MACHINE
607. Herbert Butterfield THE SCIENTIFIC REVOLUTION
629. Anne P. Carter THE ECONOMICS OF TECHNOLOGICAL CHANGE
873. Desmond King-Hele THE SHAPE OF THE EARTH

Author Index

(The authors of the articles in all five subject series are given here.)

Introduction

Once upon a time, a learned man was called a scholar or, perhaps, a philosopher. It was understood that a true scholar was learned because he had read all the available learned books written by the learned men of the past and could therefore discourse learnedly—with knowledge, and perhaps even wisdom—on all the various facets of the truths men felt they had gleaned from nature.

For most of man's history, it seemed a fair ambition, and not too grandiose, to attempt to encompass all that men had learned, so that one might say, as Francis Bacon said in 1592, "I have taken all knowledge to be my province."

At the turn of the fifteenth century Leonardo da Vinci, an inordinately gifted artist, showed himself to be just as talented an engineer, just as talented a scientist. He was the "Renaissance Man" *par excellence;* the man who was unboundedly versatile in talent and learning.

About two centuries later, Gottfried Wilhelm von Leibniz could, in a single lifetime, make first-class contributions to philosophy, theology, mathematics, history, law, and politics.

Even as late as early years of the nineteenth century, Johann Wolfgang von Goethe could find time to be not only the greatest poet the German language had ever seen, but also to make contributions of importance to comparative anatomy, conduct a geological survey, develop a theory of plant structure, and produce an ingenious—although totally wrong—theory of the nature of light. In the same period, Thomas Young could not only evolve the wave theory of light and advance theories of color vision and tidal phenomena but could also make important contributions to the decipherment of Egyptian hieroglyphics.

But through the centuries, learning had accumulated at such an ever-increasing pace that it became more and more difficult for a single mind, however brilliant, to embrace it all. The gaps left in one man's learning had, of necessity, to grow wider; the learning he actually attained had to be more superficial.

Before the middle of the nineteenth century had been reached the age of the specialist was in full flower. It became impossible for a scholar to be known as a "learned man" in both science and literature; for it to be taken for granted that a philosopher could read both Hegel and Herschel with equal understanding. So enormous had the field of science become that those scholars whose concern was with the knowledge of the past shrank away from it. The very terms "scholar" and "philosopher" were applied only to those who dealt with the humanities: with literature, art, music, and the newly emerging social studies—to John Stuart Mill, for instance. For those men who wished to learn the laws of the natural universe and to study the phenomena that included the motions of the heavenly bodies, the weathering of the rocks, the changes in living matter and its inanimate surroundings, a special term had to be invented. The term was "scientist," and Hermann von Helmholtz was an example.

It may surprise most persons to be told that although "science" is an old term, "scientist" is not. There is no record of the use of "scientist" before 1840. And even at the moment the term was invented, it was adequate only to describe a class of individuals, not the individuals themselves. No man of the time would have been so presumptuous as to identify himself with science as a whole—only with some particular branch of science. They were physicists, chemists, geologists, astronomers, biologists.

And it has got worse since then, not better. Few physicists today would be content to describe themselves as "physicists." They would be much more likely to speak of themselves as "solid-state physicists" or "nuclear physicists" or whatever sub-subclassification struck them as descriptive. A chemist might much prefer to call himself a "high-polymer chemist" rather than merely a "chemist"; a biologist, a "molecular biologist" rather than merely a "biologist" or even a "biochemist."

Indeed, when the United States government seeks to classify the nation's men of science so that in emergencies it will know to whom to turn for assistance in meeting them, literally hundreds of finely separated slivers of specialization are offered for choice. Such a monumental degree of specialization results, in part, from the number of scientists now at work: it has become a cliche to say that 90 per cent of all the scientists who ever lived are alive at this moment. And the result? In each particular sliver of science, so many are working and so much is produced that the mind and thought of each specialist is filled to bursting with the produce of his

specialty, almost inevitably to the near-exclusion of all else.

In other words, a chemist who works with elastomers finds it all he can do to keep up with advances in the chemistry and technology of elastomers. Not only does he lack time to explore science generally (let alone the humanities) but he may, perforce, be unaware of what is being done in fields closely allied to his own—in petroleum technology, in ceramics, or even in the nonelastomeric portions of the field of polymer chemistry.

Science is, in one way, a microcosm of our expanding universe, with each segment of scientific knowledge retreating steadily from every other segment. The chances of cross-fertilization become ever less as the intellectual distance between the segments increases. Esthetically, this is a sad thing—and it is a downright danger to the growth of science, too.

We know from experience that the most startling advances have their origin at the borderlines of the specialties, where the techniques developed in one field are applied, with fertile effect, to the subject matter of another. The Wilson cloud chamber, which sent nuclear physics racing forward, was the result of a thought that arose as Charles Wilson studied a meteorological observation. The metal-shadowing technique that made electron microscopy one of the most useful techniques of science was a by-product of the study of the moon's surface.

If cross-fertilization dwindles, the rate of scientific advance will almost surely dwindle as well, and so anything that encourages cross-fertilization is all to the good. In this respect, *Scientific American* magazine is unique. It is a forum where members of the scientific family can address not only their brothers, not only their distant cousins, but even the members of altogether different families. The SCIENTIFIC AMERICAN RESOURCE LIBRARY will enable these scientists to speak to an even wider audience.

Are there rarefied regions of nuclear physics? George Gamow can talk of uncertainty and of gravitation in a way that is understandable to those who are not nuclear physicists. Is cosmology a new abstract art-form on a universal scale? Martin Gardner can describe a universe in which time goes backward and Fred Hoyle can describe one which never changes, thus giving noncosmologists at least a glimpse of the cosmic glories of their starry visions.

Among the other articles in *Readings in the Physical Sciences* are many equally fascinating. Eugene M. Lifshitz describes the utterly strange events that occur in the neighborhood of absolute zero and Jesse L. Greenstein tells of the equally strange events that take place on and within a dying star. Stewart E. Miller writes about the fantastic and yet obviously practical use of lasers in long distance communication, Edwin H. Land about something as subtle as color vision and its relation to photography, and Marvin L. Minsky takes us off into the science-fictional (are we sure?) world of artificial intelligence.

We can travel from the artistry of moire patterns to the other-artistry of the water molecule, from the near-nothingness of neutrinos to the massive mystery of the quasi-stellar objects (quasars, for short), from the pion to the Pleiades, from the collapsing star to the expanding universe, from titanium to computers—

And it is all for the layman.

But who is this layman? A dictionary will tell us that he is any man not accomplished in a given art or profession, as distinguished from those who are: the physicist is a layman to the physician, the chemist to the clergyman, seismologist to the statesmen—and vice versa and ad infinitum.

Yet why should not a lawyer, an accountant, a shoe salesman, read about the muon and the Mössbauer effect? It is the interest, sympathy, and tax-money of the millions who are laymen (to scientists of whatever persuasion) that support scientific research in these days of vast government grants. It is the lives of these laymen that is bettered by science—and endangered by science, too. And then there are the countless young students who are still searching for the road that they will travel. The articles in these volumes can be for them signposts pointing down many different roads. Those who are already interested in science will read them avidly. Those who think they are *not* should also read them—not only to discover where their fellow students are going and why they are going there, but because an article on dying stars, on color vision, on the growth of crystals, on analgesic drugs, or on game-theory, may change their minds and send them down the road to a scientific career.

Although some of the articles in *Readings in the Physical Sciences* are two decades old and the information they contain is already dated, that does not mean that the articles themselves have lost their usefulness. What they contain is not wrong, it has just been superseded—and this, in itself, gives a picture of the jet-propelled advance that produces a new understanding, then leaves it behind almost in a breath.

Viewing such a picture does more than satisfy curiosity. It teaches what is perhaps the most inspiring of all truths about science—that it never reveals, but is always in the process of revealing; that it never *is*, but is always and endlessly *becoming*; and that it is thus the greatest adventure the human mind has ever known.

Isaac Asimov
Newton, Massachusetts
September, 1969

SCIENTIFIC AMERICAN Resource Library

Physical Sciences 1

SCIENTIFIC AMERICAN September 1953, Vol. 189, No. 3, pp. 58-63 OFFPRINT 201

WHAT HOLDS THE NUCLEUS TOGETHER?

by Hans A. Bethe

Electrical forces bind the electron to the atom, but they cause nuclear particles to fly apart. The powerful cohesion of protons and neutrons must be explained by a wholly different phenomenon.

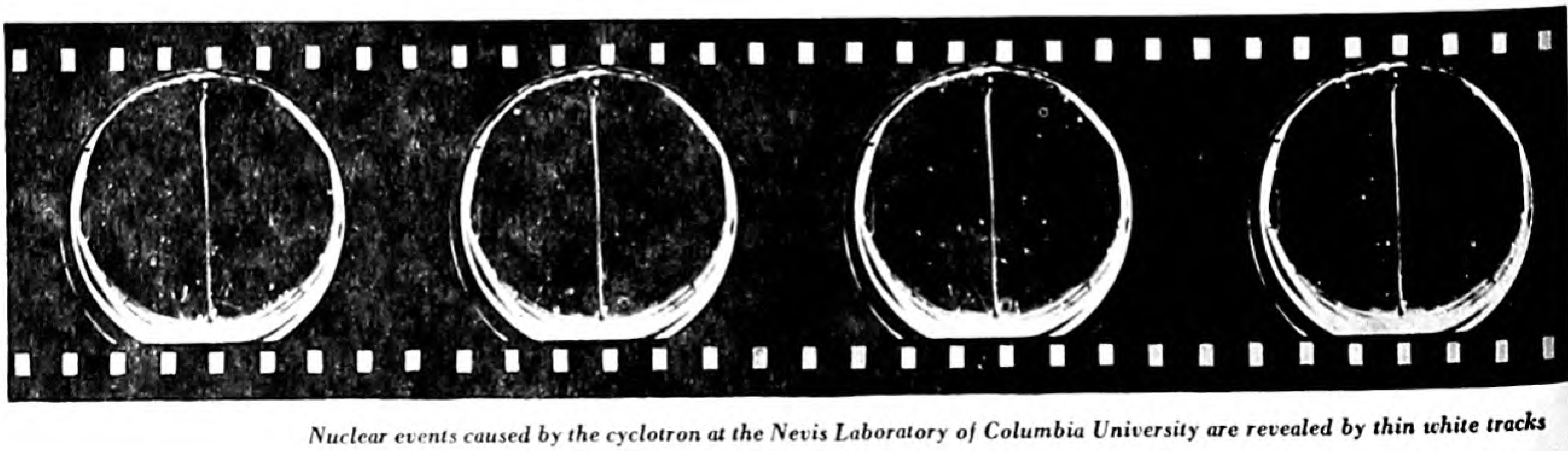
What holds the nucleus of the atom together? In the past quarter century physicists have devoted a huge amount of experimentation and mental labor to this problem—probably more man-hours than have been given to any other scientific question in the history of mankind. The problem is not only fundamental but alien to our experience. By all the laws of known forces, the particles in an atom's nucleus should flee from one another, instead of clinging together so strongly that we must build enormously energetic machines to pry them apart. The glue that holds the nucleus together must be a kind of force utterly different from any we yet know.

Let us first look briefly at the general features of the atom, which is much too small to be seen under the most powerful microscope but about which we nonetheless have a great deal of information. It is constructed of a heavy, positively charged nucleus surrounded by a "planetary system" of light, negatively charged electrons. The forces that govern the behavior of the electrons are thoroughly familiar: they are the forces of electric attraction and repulsion. To describe the motions of the electrons physicists had to invent a new mechanics known as quantum mechanics. Once this was worked out, it became possible to understand all the properties of atoms as a whole—their sizes, their chemical behavior, the light they emit, and so on—in terms of the motions of the electrons around the nucleus.

The nucleus itself is a very different problem. Its building blocks are the positively charged particles called protons and the electrically neutral particles known as neutrons. The nucleus, containing about 99.95 per cent of the total mass of the atom, is far more densely packed than the electrons in the atom's outer regions; if you were to imagine the atom as a whole to be as big as a house, the nucleus would be the size of a pinhead. Now detailed investigations of the nucleus early turned up a remarkable fact: whereas the density of the fluffy outer structure of atoms varies greatly from one kind of atom to another, all nuclei have a uniform density (about 100 trillion times that of water). Thus the total volume of an atom, insofar as its volume can be defined at all, is not necessarily proportional to its weight, a circumstance which makes some substances denser than others. But the volume of a nucleus is very nearly proportional to its weight, just as a piece of iron 10 times as heavy as another is also 10 times as large in volume.

This resemblance of nuclei to the matter of everyday experience suggested that the forces holding the nucleus together might be something like those that bind atoms together. We know that gross matter is held together by forces between neighboring atoms, and that there are no important interactions between atoms distant from one another. It is therefore assumed that the forces in the nucleus likewise act mainly between neighboring particles, rather than from one end of the nucleus to the other.

But what can these forces be? Clearly electric forces are out of the question. In the first place, the electric force between two protons is repulsive, not attractive. And even if the sign were



Nuclear events caused by the cyclotron at the Nevis Laboratory of Columbia University are revealed by thin white tracks

changed so that they attracted one another, the electric force of attraction would be too small by a factor of 40 to account for the binding energy with which protons are held together in the nucleus. Besides all this, what about the uncharged neutrons, which cannot exert any electric force, attractive or otherwise—how could the nucleus hold them?

As for gravitation, the other important force with which we are acquainted, that is completely hopeless. The gravitational force between two particles in a nucleus is too small to explain their attraction by a factor of 10^{37}!

We are confronted with a problem which is just the opposite of the one physicists had when they began to study the atom as a whole. They were completely familiar with the forces (electric) at play, but had to discover the laws (quantum mechanics) that governed the operation of these forces. In the case of the nucleus, we are fairly confident about the governing laws (again quantum mechanics), but must discover the force.

One might picture the situation in this way. You are walking in the park and come upon a group of men playing baseball. After watching for a few minutes you decide that it is a match between lunatics. The batters seem to run to any base that pleases them; the fielders throw the ball at random; the object of the game is utterly obscure, and the score, impossible to compute. But by long, intense observation you finally figure out the strange rules of the game. That is where atomic physics had arrived 20 years ago. We have now moved along to another place in the park and discovered a second game more insane than the first. The rules seem to be the same, but the players are playing without a ball! Something—we do not know just what—is passing back and forth among the players, and to understand the game we must find out what that something is. The invisible ball shuttling among the players corresponds to the force between particles in the nucleus.

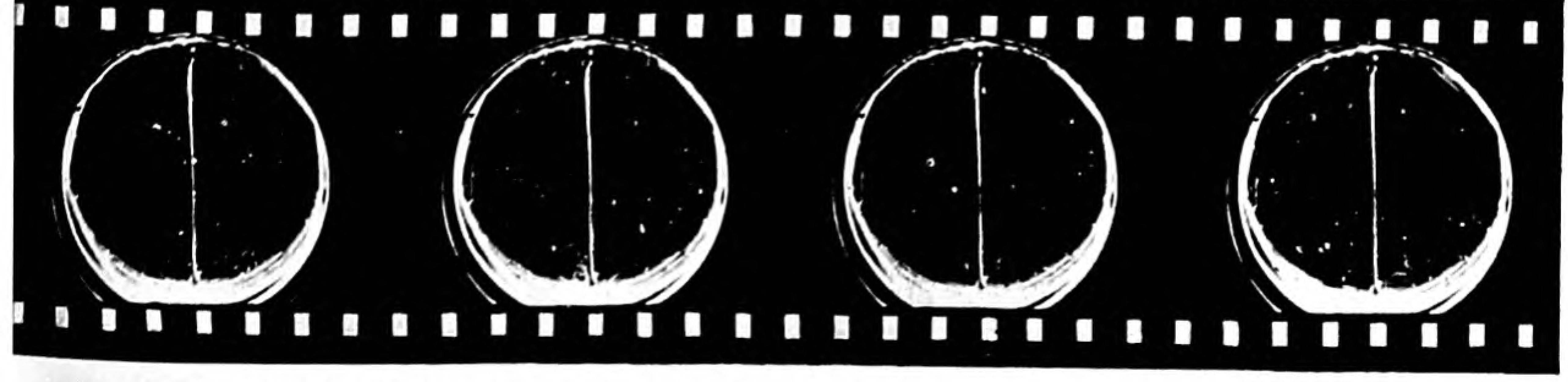
Our problem is twofold: (1) to measure the force and determine its other properties, and (2) to probe into the "cause" of the nuclear force, as it were, by studying its connections with other physical phenomena.

We can get an approximate measurement of the strength of the nuclear forces by determining the binding energy with which the nucleus is held together. This can be done in two ways: by measuring the energy set free or consumed in various nuclear reactions, or by using Einstein's relation $E=Mc^2$, which says that the binding energy is equal to the mass defect in the nucleus times the square of the velocity of light. The "mass defect," of course, refers to the fact that the mass of the nucleus is slightly smaller than the sum of the masses of the particles combined in it; the difference is the "defect."

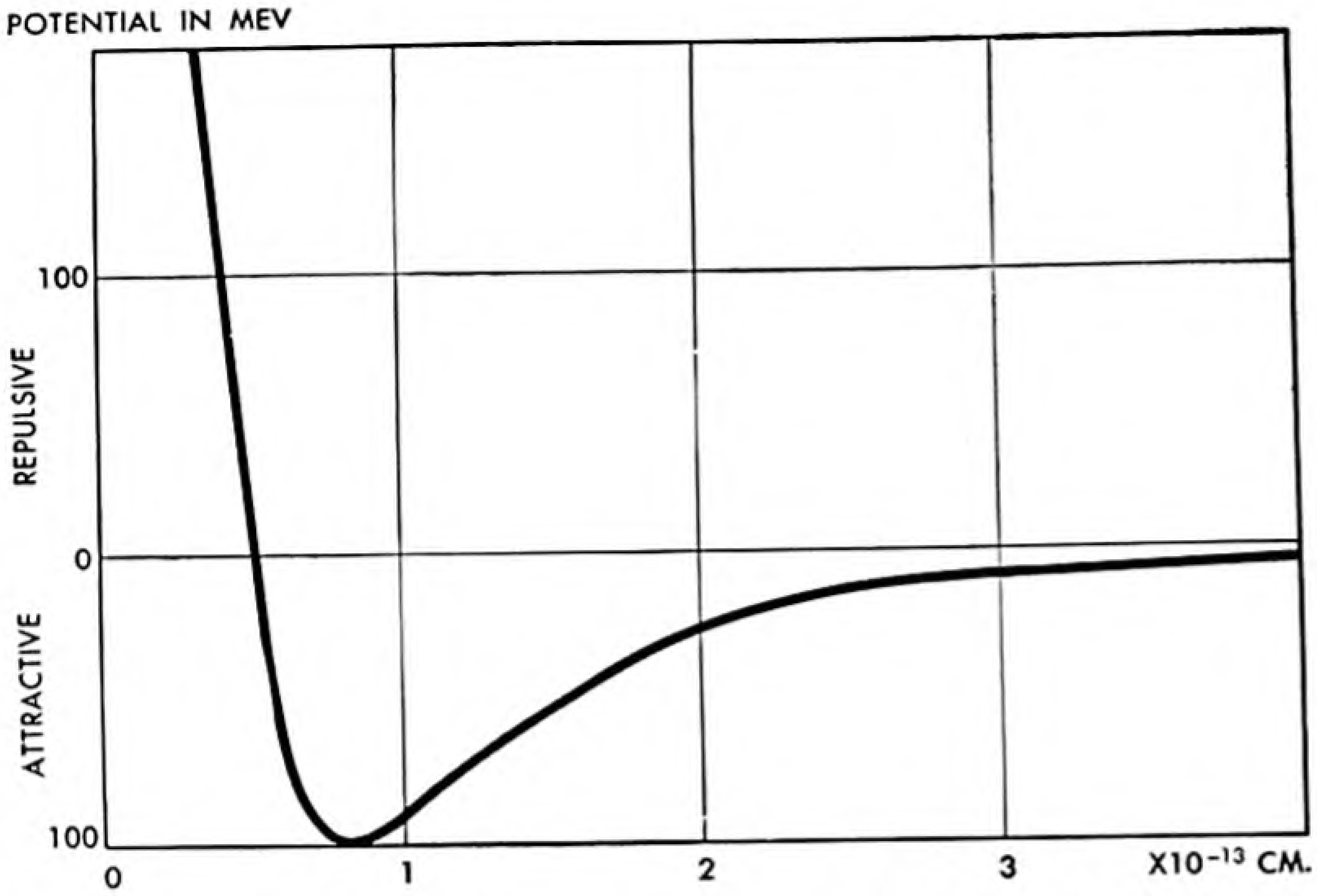
By these two methods it has been determined that the binding energy holding each particle in a heavy nucleus is between six and eight million electron volts (roughly a million times the energy that holds atoms together in a molecule). But this is still far from telling us much about the force between two individual particles—to say nothing of the complex set of interacting forces operating among the whole group of particles. We can try to use the measured energies with which particles are bound to the nucleus as a basis for calculating the nuclear force. If we tried this with any complex nucleus, however, we would be in very deep water; the mathematical problem of computing from the binding energy the forces among the 16 particles of the oxygen nucleus, for instance, is so formidable that no one would dream of attempting it. We are forced to concentrate, as the atomic physicists did in studying the atom, on the simplest possible system. The atomic physicists obtained most of their information about atoms from the two-particle hydrogen atom—one proton and one electron. As a subject for investigating nuclear forces, the simplest nucleus we can find is the deuteron (the nucleus of heavy hydrogen), which consists of one proton and one neutron.

Unfortunately the deuteron is far less helpful than the hydrogen atom was. The hydrogen atom's various energy states, normal and excited, all provide means of testing the laws governing the forces between its proton and electron. But the deuteron has no excited state. The only measurement it can give us is the binding energy in its fixed ground state, and this alone is not sufficient to determine with precision the force between its two particles and—what the nuclear physicist particularly wants to know—how that force varies with distance. We have therefore had to study the matter indirectly by investigating the interaction between free protons and free neutrons. A beam of neutrons is directed at a piece of matter containing hydrogen. Neutrons colliding with the protons in the hydrogen are scattered in various directions. By observing how many neutrons are scattered in each direction, and by using neutron beams of varying speeds, we are able to deduce the force between the proton and the neutron.

The most conspicuous feature of nuclear forces turns out to be their short range. At a distance of about 10^{-13} centimeter (which is a hundred-thousandth of the radius of an atom) the nuclear force of attraction between two protons is about 40 times as strong as the electric force of repulsion between them. At four times that distance the nuclear force has dropped off to the same strength as the electric force; at 25 times the distance the electric force is a million times stronger. On the other hand, there is some evidence that at extremely short distances (perhaps less than half of 10^{-13} cm.) the nuclear force changes from a



in a diffusion cloud chamber. A 35-millimeter camera automatically photographs the chamber once every 10 seconds



NUCLEAR FORCE (measured in millions of electron volts) is plotted against the distance between particles. When the distance is less than half of 10^{-13} centimeter, the nucleons repel one another. They most strongly attract one another at just under 10^{-13} cm.

strong attractive force to an even stronger force of repulsion.

The nuclear forces are far more complicated than electric or any other known forces. The force between two nuclear particles apparently depends not only on the distance but also on the particles' relative velocity and on the relative orientation of their spins. Moreover, there are forces which act among three, four or more particles simultaneously. Again, there is the remarkable fact that the force between particles is independent of the particles' charge. Proton and proton, neutron and neutron, proton and neutron—all have about the same attractive force toward each other. This finding is especially hard to explain because it is so contrary to the behavior of charged particles in common experience.

Another remarkable feature of the nuclear force is the kind of exchange that occurs between one particle and another. In the gross material world, and in the world of atoms, when two bodies of equal weight collide usually the faster moving body retains the greater speed or the two bodies share their energy about equally. But in the world of protons and neutrons something quite different commonly happens. When a very fast neutron hits a proton, very often the proton jumps forward with almost as much speed as the neutron had, while the neutron is stopped almost to a standstill. The simplest way to explain this is that the neutron snatches the positive charge from the proton and keeps on going, without transferring much of its momentum to the proton. In other words, the proton that suddenly jumps forward is really the original neutron transformed into a proton.

Having explored the properties of nuclear forces, we may now try to "explain" them. At this point it is appropriate to point out that for a physicist the word "explanation" has a rather different meaning from that in everyday usage. People generally explain something in terms of concepts more familiar than what they are explaining. But physicists very often "explain" a rather familiar phenomenon in terms of far less familiar concepts. For them an explanation consists in connecting different physical phenomena, building a logical structure and deriving the simplest possible mathematical formula to describe all the connected phenomena.

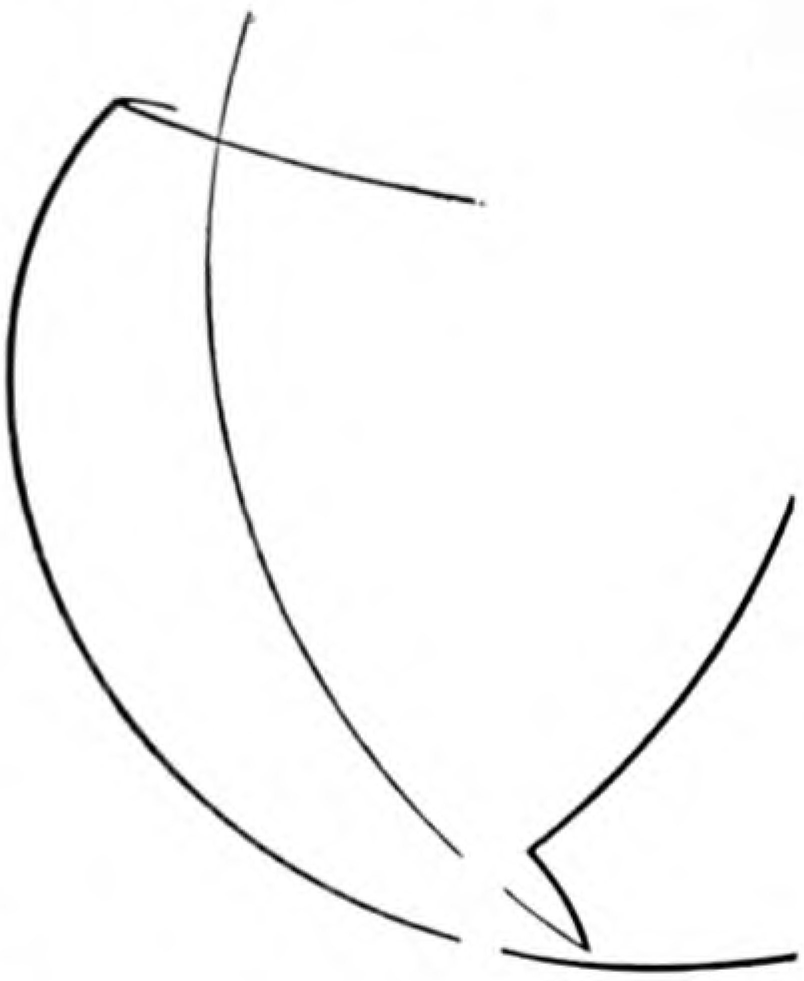
It must be clear from what has already been said that the nuclear forces cannot be explained in terms of forces with which physicists were familiar before 1930. Analogies with those forces can, however, be a starting point for a theory of the new force.

The force about which we know the most is electromagnetic force. We know that the interaction between electrically charged bodies moves with the speed of light. Further, this interaction can be described essentially by saying that quanta of light are emitted by one electric particle and absorbed by another. In the process, the light quanta transmit energy and momentum from the first to the second particle; in other words, they transmit an electric force, though they themselves have no electric charge.

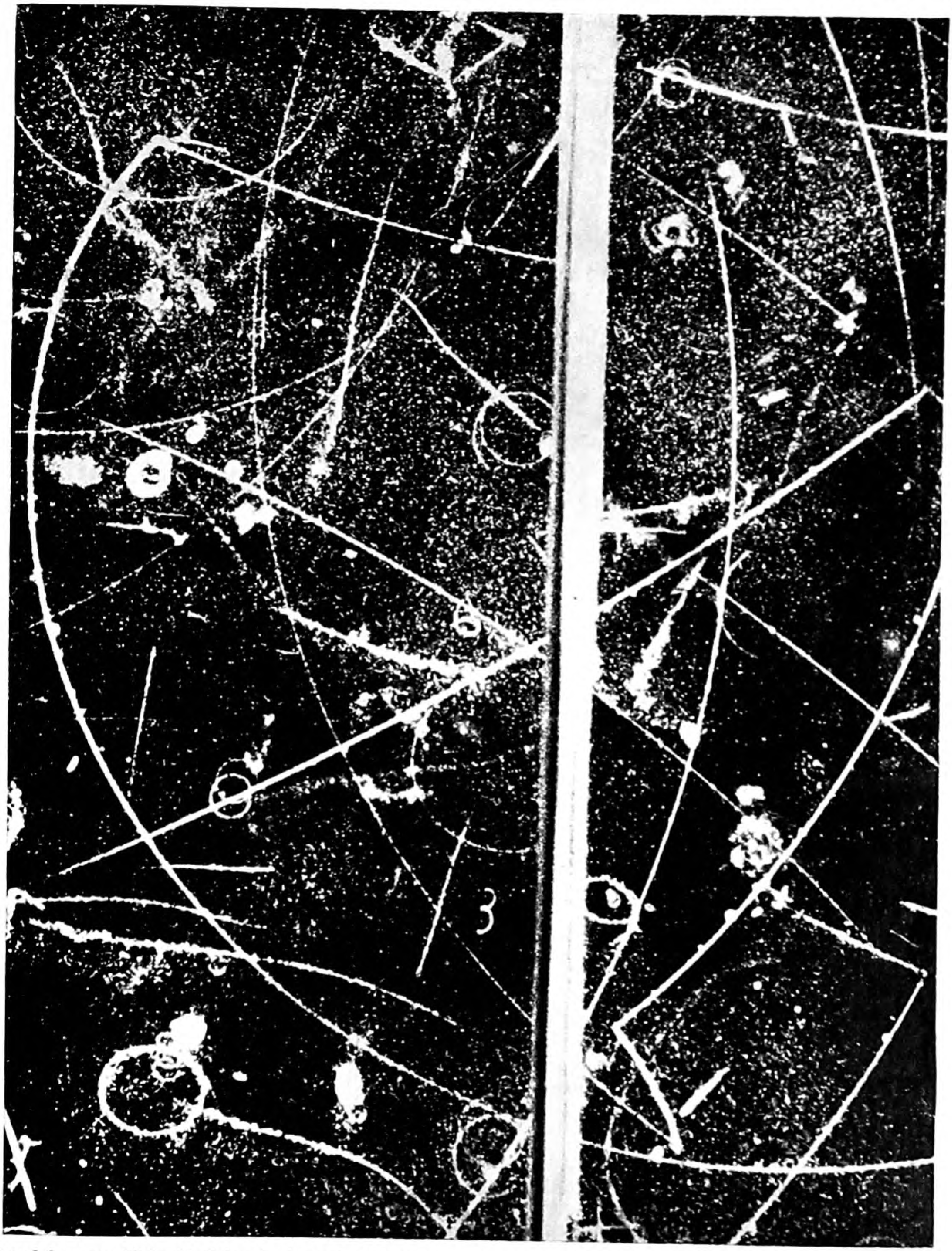
It was natural to assume that nuclear forces behave like electromagnetic ones, and this suggestion was made by the Japanese physicist Hideki Yukawa as early as 1935. In Yukawa's theory, in the nucleus the role of the light quantum is taken by a new particle, whose emission and absorption is supposed to transmit the nuclear forces. This particle, when Yukawa invented it, was of course purely hypothetical. Todav it is known as the meson.

Yukawa next tried to figure out what properties his hypothetical particle should have. First of all, he noted that the short range of nuclear forces would be explained if the mesons were supposed to have a mass—in contrast to light quanta which have none. In fact, he worked out the range of the nuclear forces mathematically in terms of Planck's constant, the velocity of light and the mass of the meson. He estimated that the meson mass should be between 100 and 200 times the mass of the electron. (Today we know that 300 is a better figure.)

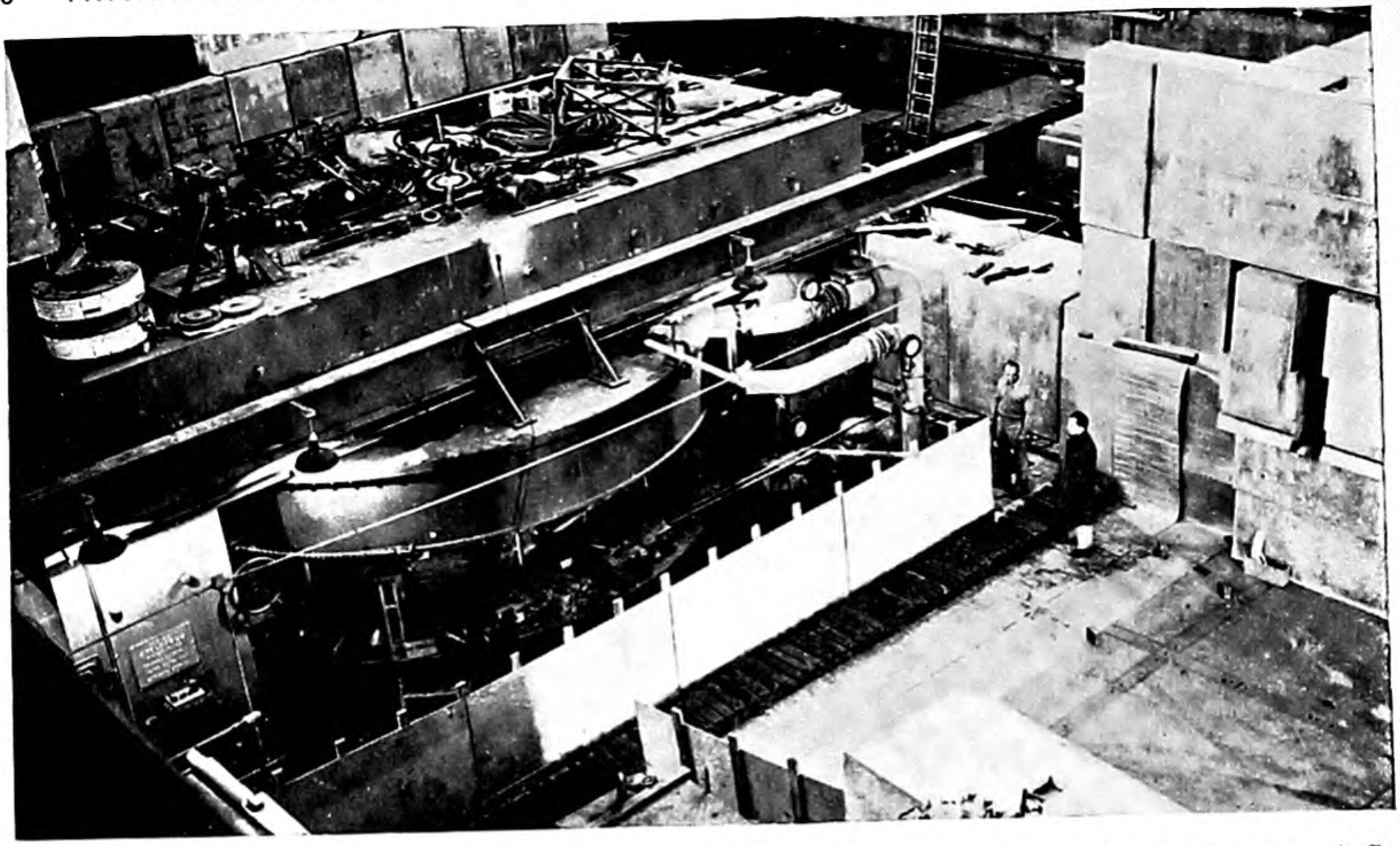
Secondly, Yukawa suggested that to explain exchange forces the mesons must be charged. When a proton and a neutron interact, he postulated, the proton may emit a positive meson which is absorbed by the neutron. In this process the



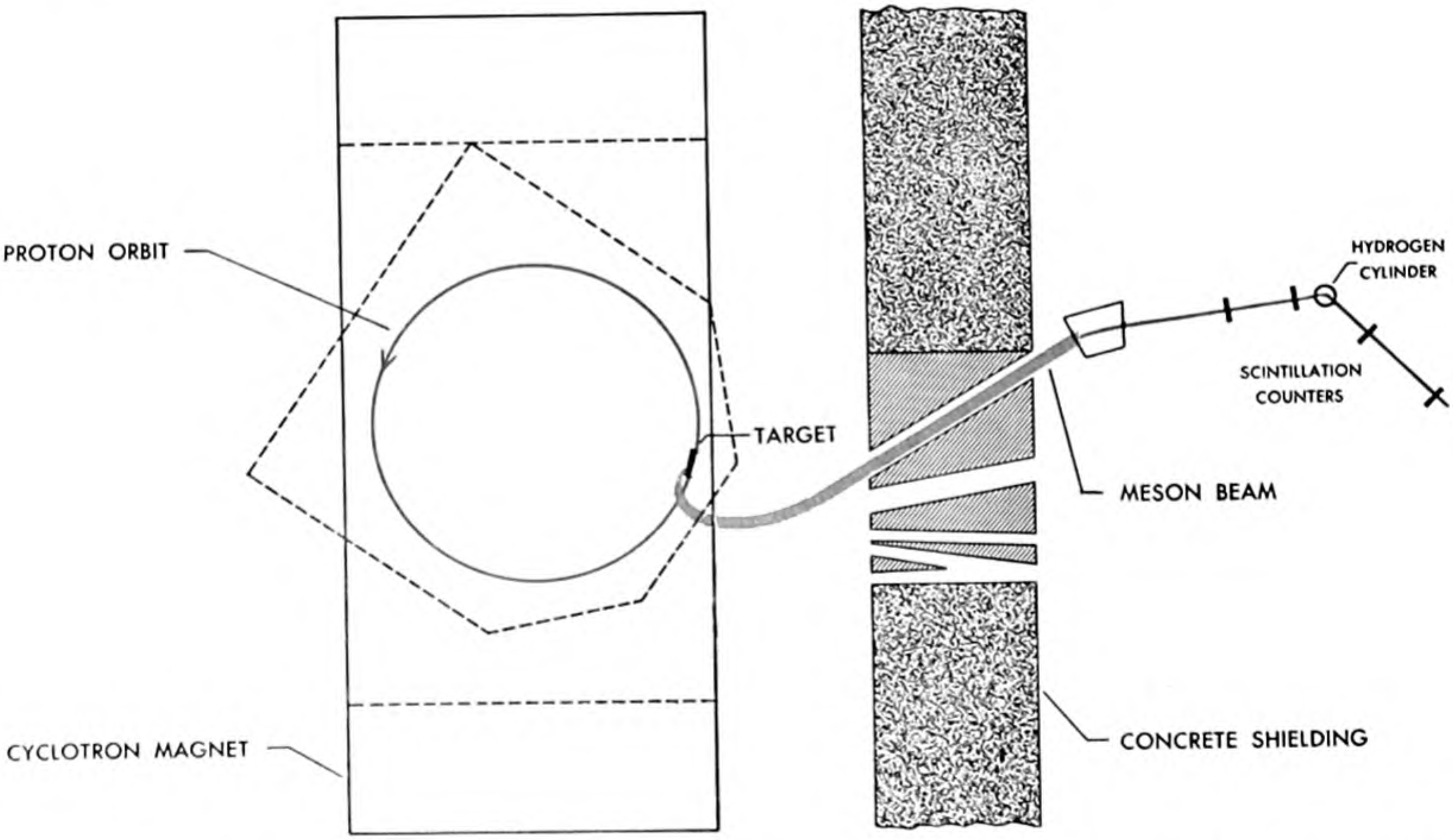
TWO MESON "EVENTS," which are diagrammed above, can be seen in a cloud chamber photograph from the Nevis Cyclo-



tron Laboratory of Columbia University. A negative pi meson entering at bottom right splits into a high-energy pair of electrons, positive and negative, at top left. Another negative pi meson entering at right center decays into a mu meson (first jog) and then into a high-speed electron (second jog). The object running from top to bottom of the picture is not in the path of the particles.



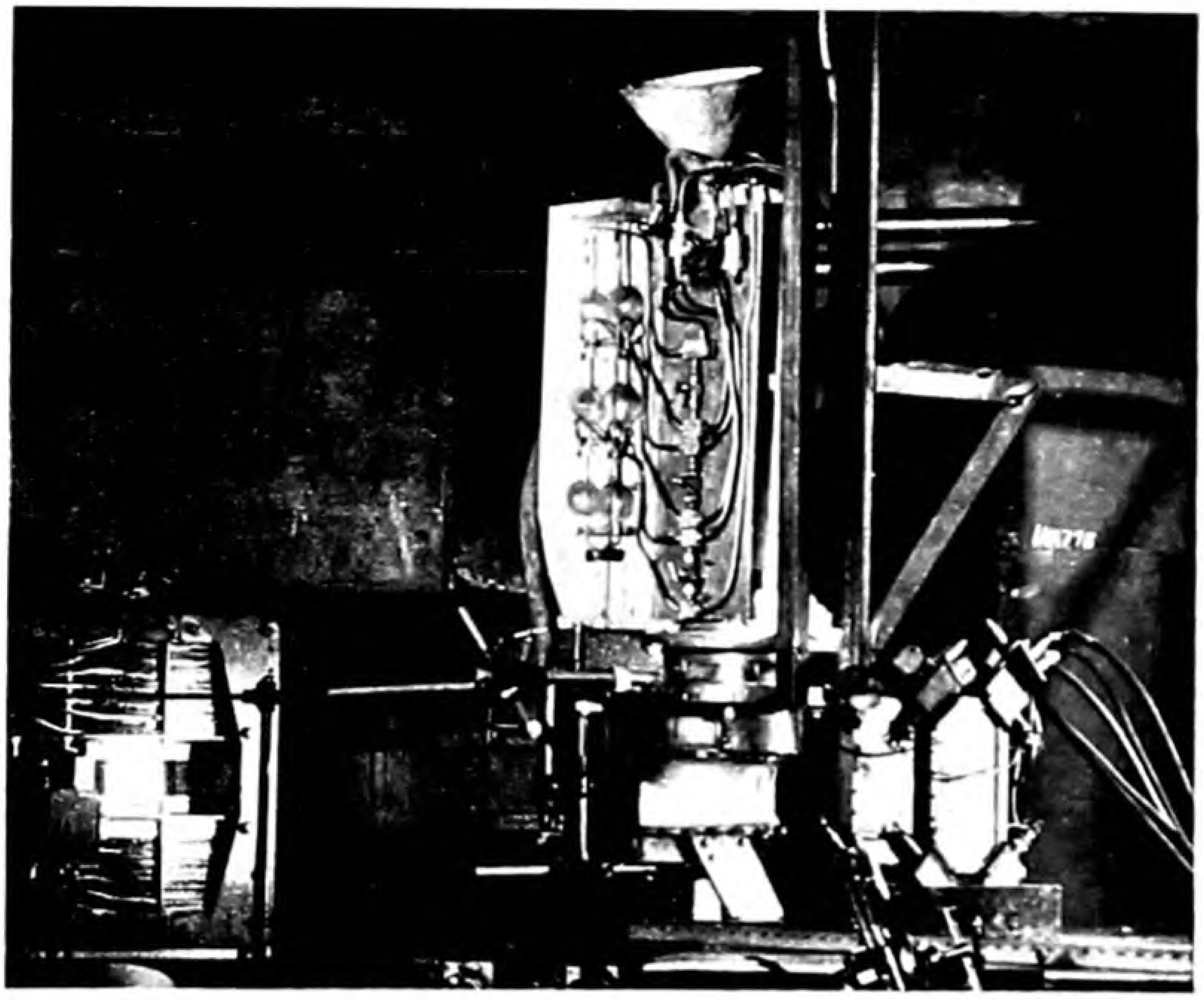
CYCLOTRON at Nevis Laboratory operates at 385 million electron volts and can produce mesons. In this picture the six-foot-thick concrete shielding blocks have been removed to the pile at right, exposing the circular magnetic pole pieces (*center*). Between the magnets can be seen the vacuum chamber, which connects with the pumping system through the system of pipes to the right.



MESON SCATTERING experiment is shown in outline. Accelerated protons (*red circle*) hit metal target to form mesons (*red line*) which are deflected by magnet through a port in iron block (*cross-hatched*). Second magnet to right of port steers them to hydrogen sample which scatters them. Counters placed ahead of sample and behind it insure that only scattered mesons are detected.

proton loses its positive charge and becomes a neutron, while the neutron gains a unit of positive charge and turns into a proton. The same result is obtained, of course, if the neutron emits a negative meson which is absorbed by the proton. Yukawa suggested the existence of both positive and negative mesons, in conformity with a general principle of physics: that for every positively charged particle there is a negatively charged counterpart. The best known examples are the negative and positive electrons.

Three years after Yukawa proposed the meson, physicists found in cosmic radiation a particle which seemed to have just the properties he had predicted. It had a mass about 200 times that of the electron and was found in both positive and negative forms. But this particle turned out to be the wrong answer—it did not interact strongly with nucleons and therefore could not transmit nuclear forces. At length in 1947 C. F. Powell, G. P. S. Occhialini and C. M. G. Lattes—an Englishman, an Italian and a Brazilian working together—discovered another particle which *did* interact strongly with nucleons and had a mass of 276 electron masses. There is every indication that this is Yukawa's meson. It is known as the pi meson, or "pion." (If kept away from nucleons it will decay after a moderately short time into one of the earlier discovered mesons, now called mu mesons.) Since the discovery of the pion, heavier mesons have been found, but probably they are less important for nuclear forces.



SCATTERING EXPERIMENT setup is photographed above. At left is port through which mesons emerge. Large apparatus topped by funnel supplies liquid hydrogen to aluminum chamber, where scattering occurs. The rectangular objects at right are two of the counters.

The next step in the theory was taken by the English physicist N. Kemmer. He reasoned that there should be a neutral meson, in order to explain the interaction between proton and proton (or neutron and neutron). A proton cannot absorb a positive meson, for it cannot acquire a second positive charge. Therefore no single charged meson could transmit a force between protons (though the simultaneous exchange of two mesons of opposite charge might, it is true, do so). Kemmer consequently suggested that a neutral meson might carry the forces between proton and proton, or, for that matter, between unlike nucleons. His theory accounted for the nuclear forces' independence of charge.

Soon after the pions were discovered in cosmic radiation, it became possible to produce them artificially with large new cyclotrons. Physicists now could obtain mesons in large quantities and explore their properties and interactions with nucleons. They soon confirmed the existence of Kemmer's neutral mesons. As for the interaction of mesons and nucleons, exact calculations will probably remain very difficult for a long time—incomparably more difficult than the calculation of electric and atomic phenomena. The main reason is that the interaction between a meson and a proton or neutron is exceedingly strong—about 1,000 times stronger than that between an electron and the electric field. The mathematical methods of quantum theory are all adapted to the weak interactions of electrodynamics.

Once the interaction between mesons and nucleons has been worked out, one can then try to derive that between two nucleons. As we have seen, the mass and charge of mesons are sufficient in themselves to explain the nuclear forces' short range, their exchange property and their independence of charge. Other aspects of meson theory can account for the dependence of nuclear forces on the direction of the spins of the nucleons and of the line joining their positions, for the strong repulsion between nucleons at extremely small distances and for the simultaneous interaction between more than two nucleons.

In short, the meson theory already accounts for all the qualitative features of nuclear forces. It looks as if we have found the ball with which the nuclear game is played. But we cannot be sure until we have figured out whether our theory can explain the behavior of the participants in quantitative terms that is, whether the range of forces, the strength of interactions and other quantities derived from the theory by calculation are of the right magnitude.

A promising start has been made by the French physicist Maurice Lévy, working at the Institute for Advanced Study in Princeton. He has shown that calculations based on the observed mass of the meson do indeed yield the correct figure for the range of the nuclear forces. His work, in combination with the theoretical work of others, also shows that the strength of interaction between nucleon and meson required to explain nuclear forces is about the same as that required to explain the scattering of mesons by protons. *About* the same—actually, the two numbers differ by approximately 50 per cent, but probably this difference is simply a measure of our mathematical ineptness in dealing with large forces. Thus the indications are that physicists are on the right track in explaining nuclear forces by transfer of mesons. But it will be a long time before our mathematical tools are developed sufficiently to determine whether the meson theory *really* explains the forces in all details.

The Author

HANS A. BETHE is a professor of physics at Cornell University. He was born in Strasbourg in 1906. His mother and grandmother were daughters of university professors, and his father was an eminent physiologist. Bethe was educated at the Universities of Frankfurt and Munich, receiving his Ph.D. from the latter institution in 1928. For two years he worked under Ernest Rutherford at Cambridge University and Enrico Fermi in Rome on a fellowship from the Rockefeller International Education Board. He left Germany for the British Isles when the Nazis came to power. After two years in England he came to the U.S. to join the Cornell faculty in 1935. In 1938 he attended a conference in theoretical physics at which the problem of the source of the sun's energy was discussed. Six weeks later he had worked out the famous carbon cycle of thermonuclear reactions, which is his best known work. During the war Bethe served first on the staff of the Radiation Laboratory at the Massachusetts Institute of Technology and then as chief of the theoretical physics division of the atomic bomb project at Los Alamos.

Bibliography

ELEMENTARY NUCLEAR THEORY. H. A. Bethe. John Wiley & Sons, Inc., 1947.

THEORETICAL NUCLEAR PHYSICS. John M. Blatt and Victor F. Weisskopf. John Wiley & Sons, Inc., 1952.

SCIENTIFIC AMERICAN April 1958, Vol. 198, No. 4, pp. 34-39 OFFPRINT 202

ANTI-MATTER

by Geoffrey Burbidge and Fred Hoyle

The discovery of the anti-proton and the anti-neutron raises the following question: Do anti-atoms made of these and other anti-particles exist in the universe?

Ever since the discovery of the positron—the opposite number of the electron—physicists have speculated about the possible existence of anti-matter. If an anti-electron could exist, why not an anti-proton and an anti-neutron? Within the last three years the Bevatron, the great accelerator at the University of California, has indeed produced anti-protons and anti-neutrons [see "The Antiproton," by Emilio Segrè and Clyde E. Wiegand; SCIENTIFIC AMERICAN Offprint 244]. Since electrons, protons and neutrons are the basic building stones of atoms, we have strong grounds for asking: If anti-particles, why not anti-atoms—that is to say, anti-matter which is the symmetrical opposite of the matter we know?

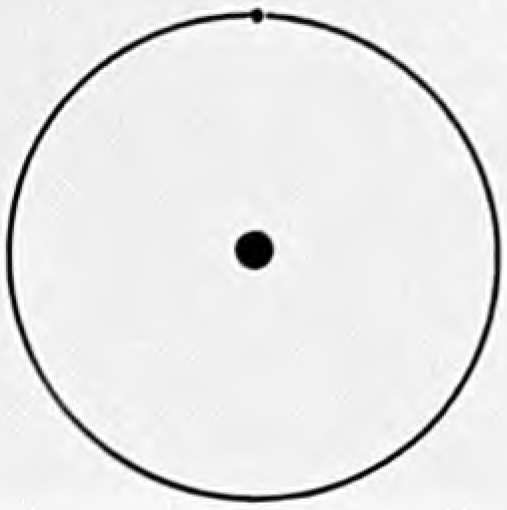
There can be no anti-matter on the earth, because particles and anti-particles annihilate each other the instant they meet. If there is anti-matter in space, we cannot recognize it with our telescopes, for anti-matter should look exactly like ordinary matter. But the problem of looking for evidence of anti-matter in the universe is not entirely hopeless. We shall consider here some possible indirect evidence on the subject. Such inquiries are not altogether academic, for, although the question of the existence of anti-matter has no practical importance to us on the earth, it does raise fundamentally important questions in modern physics and cosmology.

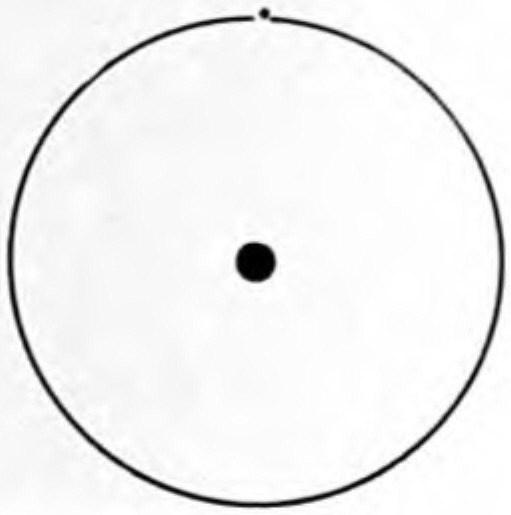
Our starting point is the known fact, established by laboratory experiments, that when a particle and an anti-particle collide they destroy each other, converting the entire mass of both particles into energy. The energy released by such an annihilation amounts to two times mc^2, in accordance with Einstein's famous equation for the conversion of mass into energy. The mutual annihilation of a proton and an anti-proton meeting at low velocity, for example, releases about 1.8 billion electron volts of energy. This energy emerges first in the form of mesons, but the mesons speedily decay and the ultimate carriers of the energy are gamma rays, neutrinos and very high-speed electrons and positrons [*see diagram on page 11*].

If we could find evidence of such annihilations going on somewhere in the universe, we might have a proof of the existence of anti-matter. Let us examine our own galaxy, the Milky Way, and start with the gas in its interstellar space. Assume that this thinly dispersed gas, made up mainly of hydrogen at an average density of only one atom per cubic centimeter, contains some anti-protons. Would the annihilating collisions between these anti-protons and protons produce any observable effects on the interstellar gas?

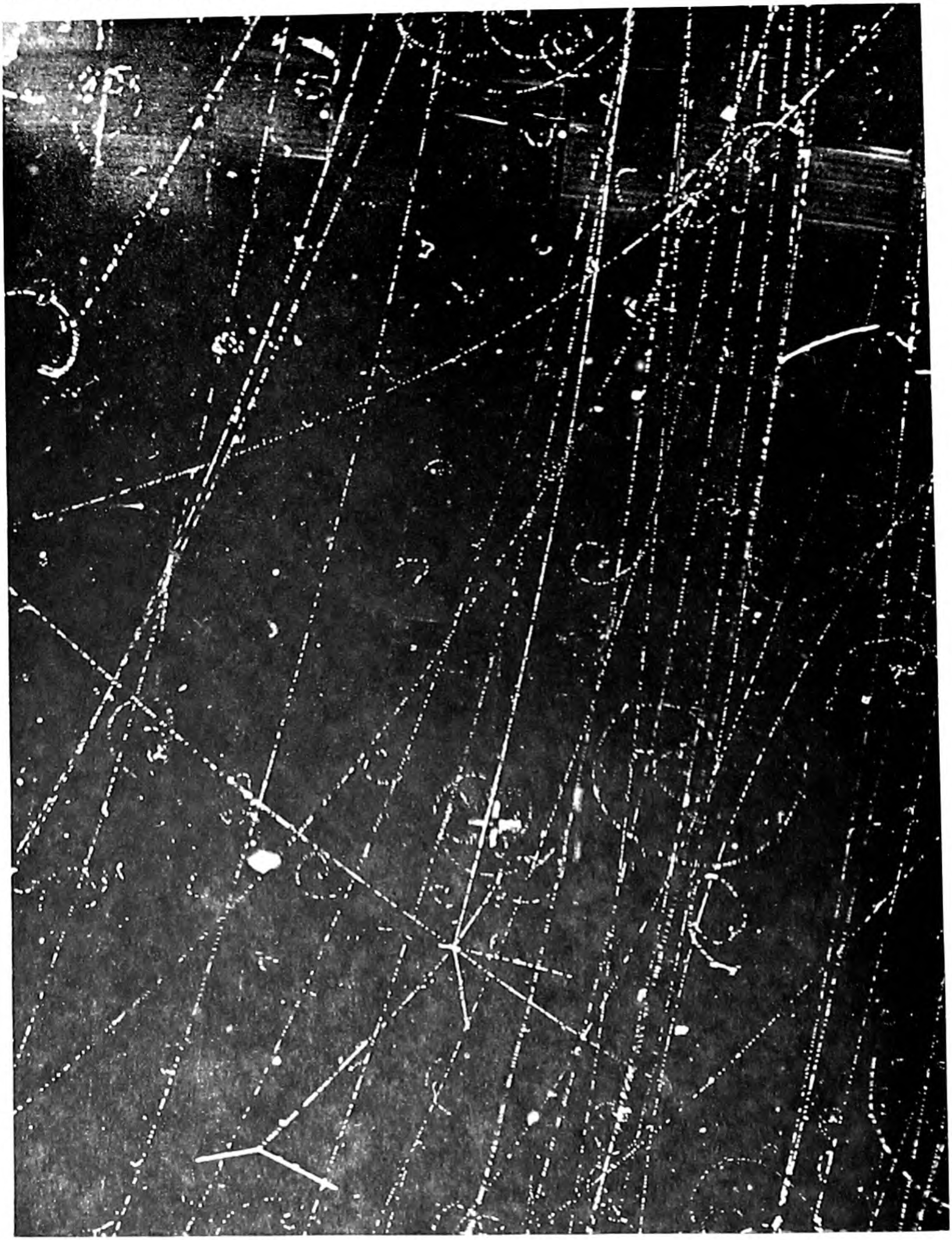
About 90 per cent of the energy generated in an annihilation is carried by gamma rays and neutrinos. Almost all of this energy would escape from our galaxy into outer space, for the chances of gamma rays or neutrinos being intercepted and absorbed by the atoms in the thin gas are exceedingly small. The situation is otherwise, however, for the electrons and positrons that carry the remaining 10 per cent of the energy from annihilations. These charged particles would be trapped inside our galaxy by its magnetic field: there is considerable evidence that our galaxy does possess a weak magnetic field.

Now the high-energy electrons and positrons would gradually give up their energy to the gas as a whole—mainly by exciting and ionizing the atoms and by electron-positron annihilations. This injection of energy would have the effect of heating the gas and generating turbulent motions. We know that there are other processes which heat and stir up the interstellar gas in our galaxy. But suppose we assume, for the sake of putting an outside limit on the total possible amount of anti-matter in our galaxy, that all the energy of the gas is generated by annihilation of anti-matter. We know from various observations what the total energy of the gas in the galaxy is—it amounts to about one 100-billionth of an erg per cubic centimeter. On this basis we can calculate that the ratio of anti-matter to ordinary matter in our





ANTI-HYDROGEN ATOM (*right*) would have an anti-proton (*black circle*) and positron (*colored dot*). Hydrogen atom (*left*) has a proton (*colored circle*) and electron (*black dot*).

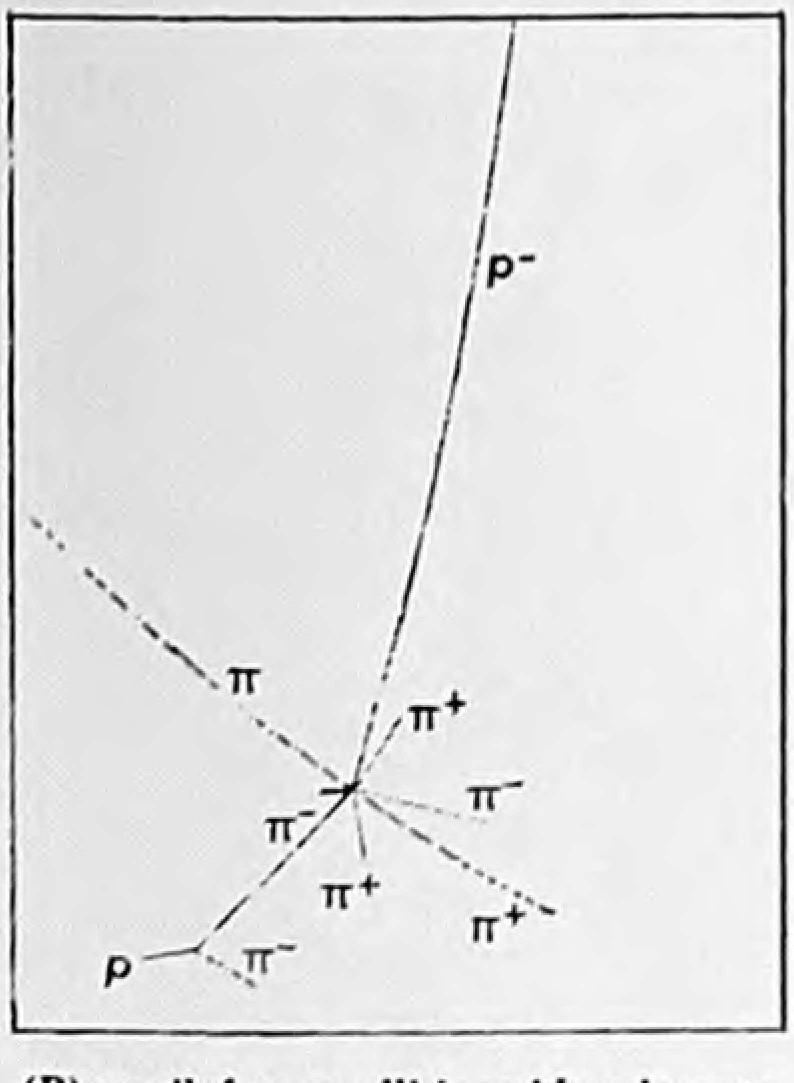


ANNIHILATION OF AN ANTI-PROTON made by the Bevatron at the University of California is recorded by these tracks in a liquid-propane bubble chamber. The event is outlined in the drawing at right. The anti-proton is annihilated by an encounter with a proton in a carbon nucleus. The short track to the left of the "star" of pi mesons is a fragment of the nucleus. At lower left a proton

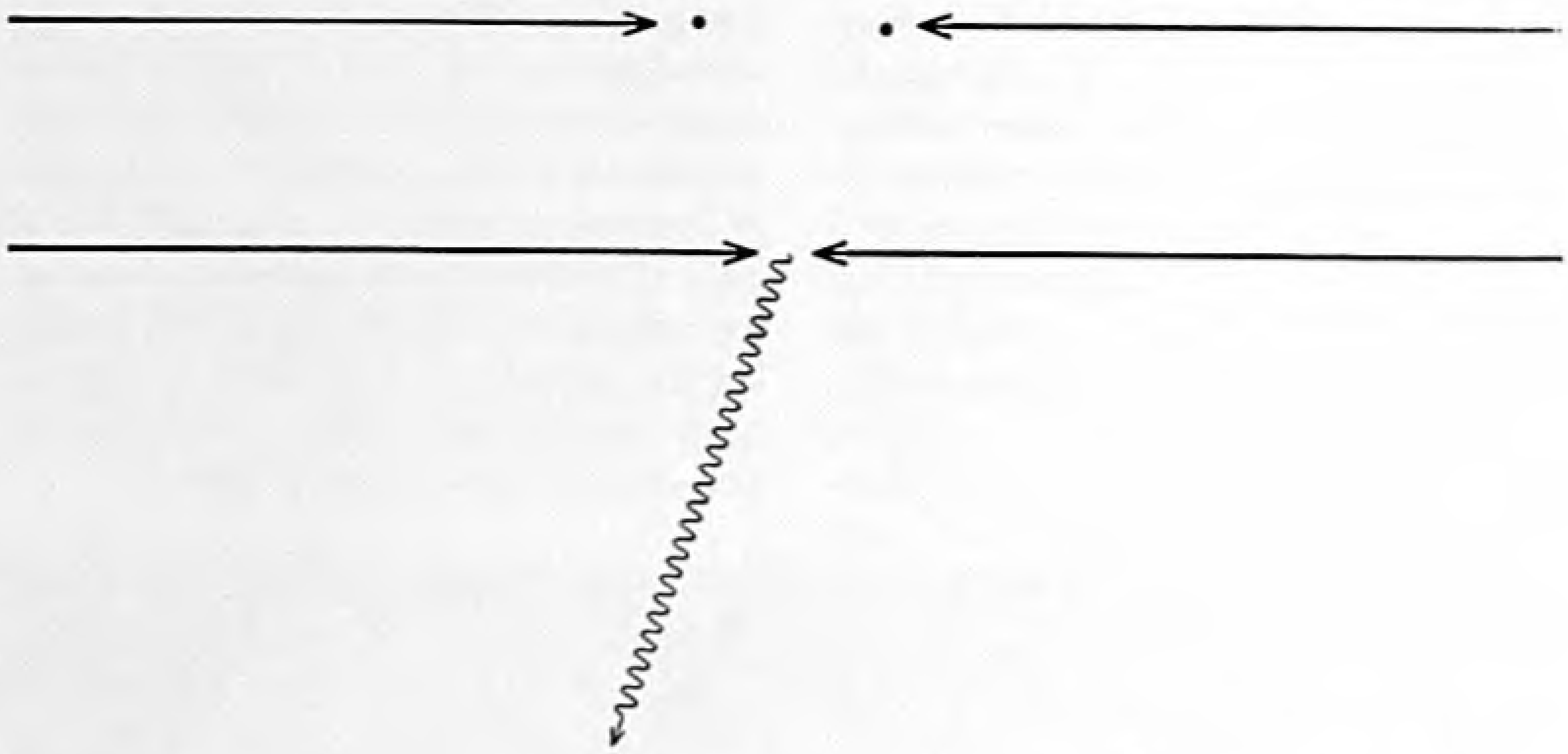
interstellar gas cannot possibly be more than one part in 10 million, spread thinly through the galaxy. From this small proportion we must deduce, incidentally, that even if anti-matter can somehow segregate itself from ordinary matter to form stars, it is extremely unlikely that there are any stars of anti-matter in our galaxy.

The establishment of a maximum figure for the amount of anti-matter does not prove that it is present, but it enables us to go on to an interesting speculation. This has to do with the recently discovered radio waves in space, stemming from so-called radio "stars" and from our galaxy as a whole. We know that electrons and positrons accelerated by a magnetic field emit a type of radiation called synchrotron radiation. This radiation can take the form of radio waves. The question then arises: Is annihilation of anti-matter responsible for some of the mysterious radio broadcasts we are receiving from space?

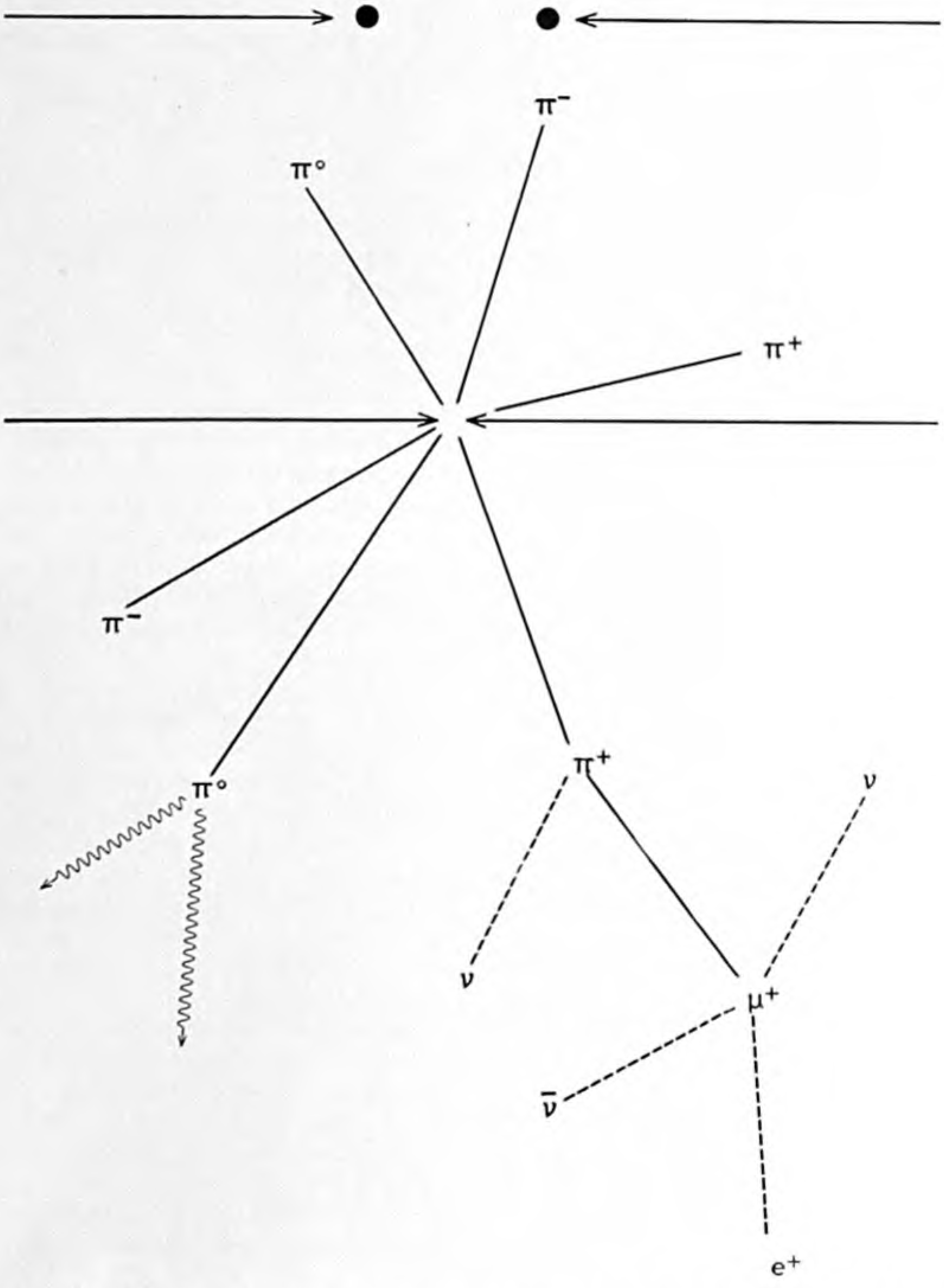
A particularly good subject for examining this question is the famous Crab Nebula, one of the strongest radio emitters in our galaxy. There is persuasive evidence that the radio emission and most of the light from this remnant of a supernova are synchrotron radiation [see "The Crab Nebula," by Jan H. Oort; SCIENTIFIC AMERICAN, March, 1957]. The Nebula appears to have a comparatively strong magnetic field and particles traveling at extremely high speeds. It is therefore conceivable that the radio emission of the Crab Nebula arises from



(P) recoils from a collision with a pi meson. The experiment was performed by groups under W. M. Powell and Emilio Segrè.



ELECTRON-POSITRON annihilation occurs when particle and anti-particle collide. If the annihilation occurs in the field of an atom, the mass of the particles may be converted into only one photon (*wavy line*). If it occurs in free space, two photons will be emitted.



PROTON–ANTI-PROTON annihilation converts the particles to pi mesons (π), which decay as indicated at bottom. In this purely schematic diagram the Greek letter ν represents a neutrino; the same letter with a line over it, an anti-neutrino; the letter μ, a mu meson.

electrons and positrons which have been created by annihilation of anti-matter. Calculating how much radio energy would be generated if one part in 10 million of the interstellar gas were anti-matter, we arrive at a figure which is close to the actual radio output of the Nebula (about 10^{33} ergs per second). The synchrotron radiation at visible wavelengths may arise from acceleration of some of the electrons and positrons to very high speeds by fast-moving gas clouds in the Nebula.

This is still no proof of the existence of anti-matter, because the original explosion of the supernova might account for its high-energy electrons and positrons. All that can be said is that the items of evidence we have considered are consistent with the possible presence of some anti-matter in our galaxy.

We come now to the wider scene. Are there, outside our own galactic system, galaxies entirely composed of anti-matter? If there are, we might possibly detect their existence if we saw an ordinary galaxy and an anti-galaxy in collision: this should be a really violent event. Here again we have one or two interesting cases for study.

Some astronomers believe that the extraordinary object called Cygnus A is a pair of galaxies in collision [see "Colliding Galaxies," by Rudolph Minkowski; SCIENTIFIC AMERICAN, September, 1956]. We are getting exceptionally strong radio signals from this object, even though it is very far away—at least 270 million light-years. The two colliding galaxies might be systems of matter and anti-matter, but suppose we make the less radical assumption that both consist predominantly of ordinary matter and each contains one part in 10 million of anti-matter. Assuming further that the magnetic fields of the two galaxies have combined to accelerate electrons and positrons, we can calculate that the annihilation of anti-matter in the colliding galaxies would generate a total of about 10^{44} ergs per second of radio energy. According to the measurements of radio astronomers, Cygnus A is actually emitting radio energy at precisely this rate!

The case of the galaxy known as Messier 87, another strong radio emitter, is even more interesting—indeed, it was M 87 that first aroused the speculation that a galaxy might contain anti-matter. M 87 looks like an unusually bright but normally shaped galaxy; there is no evidence that it represents a pair of galaxies in collision. However, it is emitting very powerful synchrotron radiation at radio wavelengths and has a bright jet or streak emitting such radiation at wavelengths of visible light [*see photograph at bottom of page 14*]. Astrophysicists have been at a loss to account for its extremely strong radiation, and it is tempting to suppose that the energy is coming from the galaxy's capture of a gob of anti-matter from an anti-galaxy.

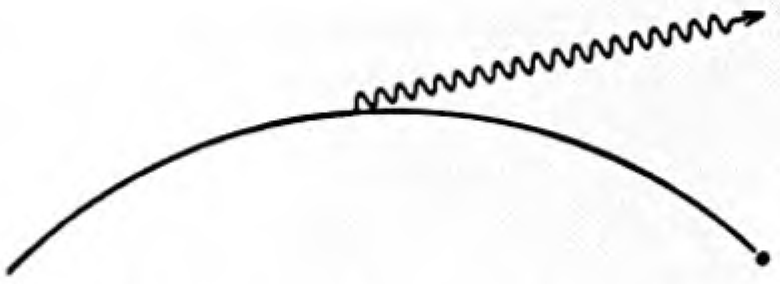
Yet notwithstanding the evidence tending to uphold the idea of the existence of anti-matter on the cosmic scale, most physicists, including the authors of this article, do not look too favorably on the hypothesis. Let us consider some of the difficulties.

In all present theories of the history of the universe—evolutionary or steady-state—symmetry arguments demand that if anti-matter exists both matter and anti-matter must be created in equal amounts. The evolutionary theory would require that the original nucleus from which the universe expanded must have contained both kinds of matter in equal parts; the steady-state hypothesis of the continuous creation of matter would imply that matter comes into being as pairs of particles and anti-particles. In either case, atoms and anti-atoms must somehow be separated selectively if they are to condense into stars and galaxies—otherwise they would destroy each other. It appears that the only way they could be so separated is by a gravitational force of repulsion between atoms and anti-atoms—in short, by anti-gravity, as opposed to the gravitational attraction that operates between atom and atom of ordinary matter.

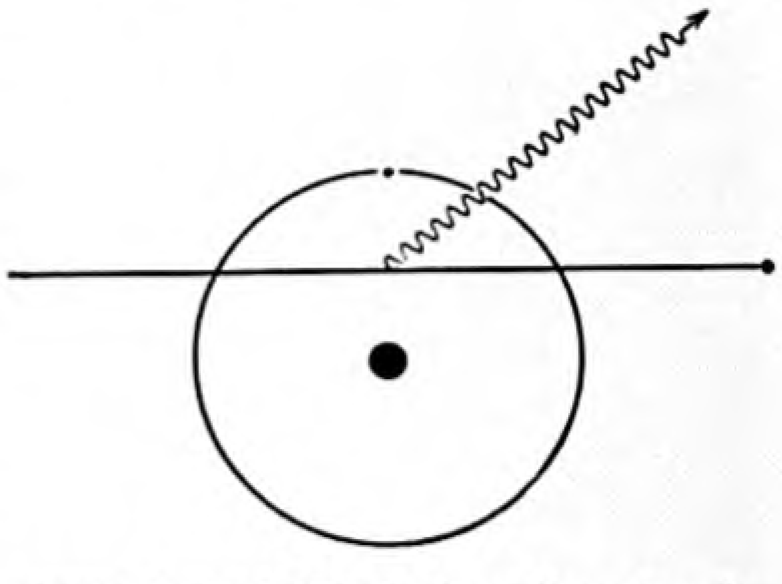
It is upon this rock that the anti-matter ideas have foundered. For the idea of anti-gravity cannot be accepted without destroying basic principles of the general theory of relativity. The successes of the relativity theory are so great that most scientists are not prepared at the present time to consider with equanimity the very considerable upheaval that would come about if it had to be abandoned or drastically modified.

Experiments designed to look for the existence of anti-gravity are possible in principle and may be worth doing. One obvious test would be to generate a beam of anti-protons in an accelerator and project it over a path parallel to the earth to see whether it would rise or fall; if the beam rose, it would indicate that anti-gravity was operating.

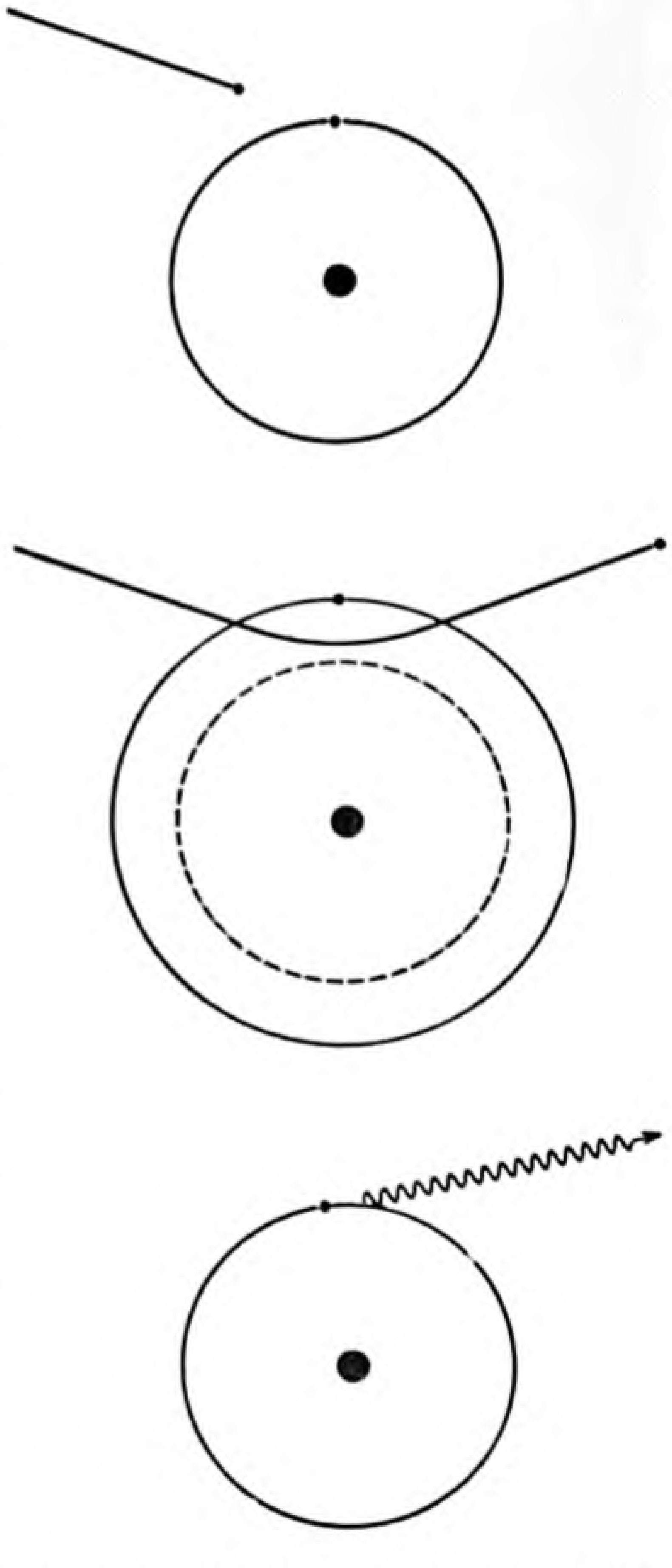
Maurice Goldhaber of the Brookhaven National Laboratory has speculated on the possible existence of two separate



RADIATION from electron in curved path transfers energy to the surrounding gas.



ENERGY TRANSFER also takes place by means of radiation emitted when an electron passes through the field of an atom.



EXCITATION of atomic electron to higher energy level is a third way to take energy from free electrons. When the excited electron drops back to orbit, it radiates energy.

worlds—one composed of matter and the other of anti-matter. Inspired by the primeval atom of Abbé Georges Lemaitre, he suggests that the universe originated from a single "particle" called the "universon." This divided immediately into a pair of "particles"—the "cosmon" and the "anti-cosmon." They flew apart with great kinetic energy (by some unspecified process) and eventually decayed, one giving rise to the cosmos we know, the other to an anti-cosmos beyond reach of our observation. Goldhaber goes on to speculate on whether some anti-matter from the anti-cosmos may have been injected into our cosmos; possibly this would be the source of the radio energy being emitted by some of our galaxies and gas clouds.

Some cosmologists have become more receptive to the idea of two separate universes since the recent overthrow of the parity principle—that is, the discovery that certain particles of our universe are unsymmetrical, in the sense that they have a particular "handedness." It be-

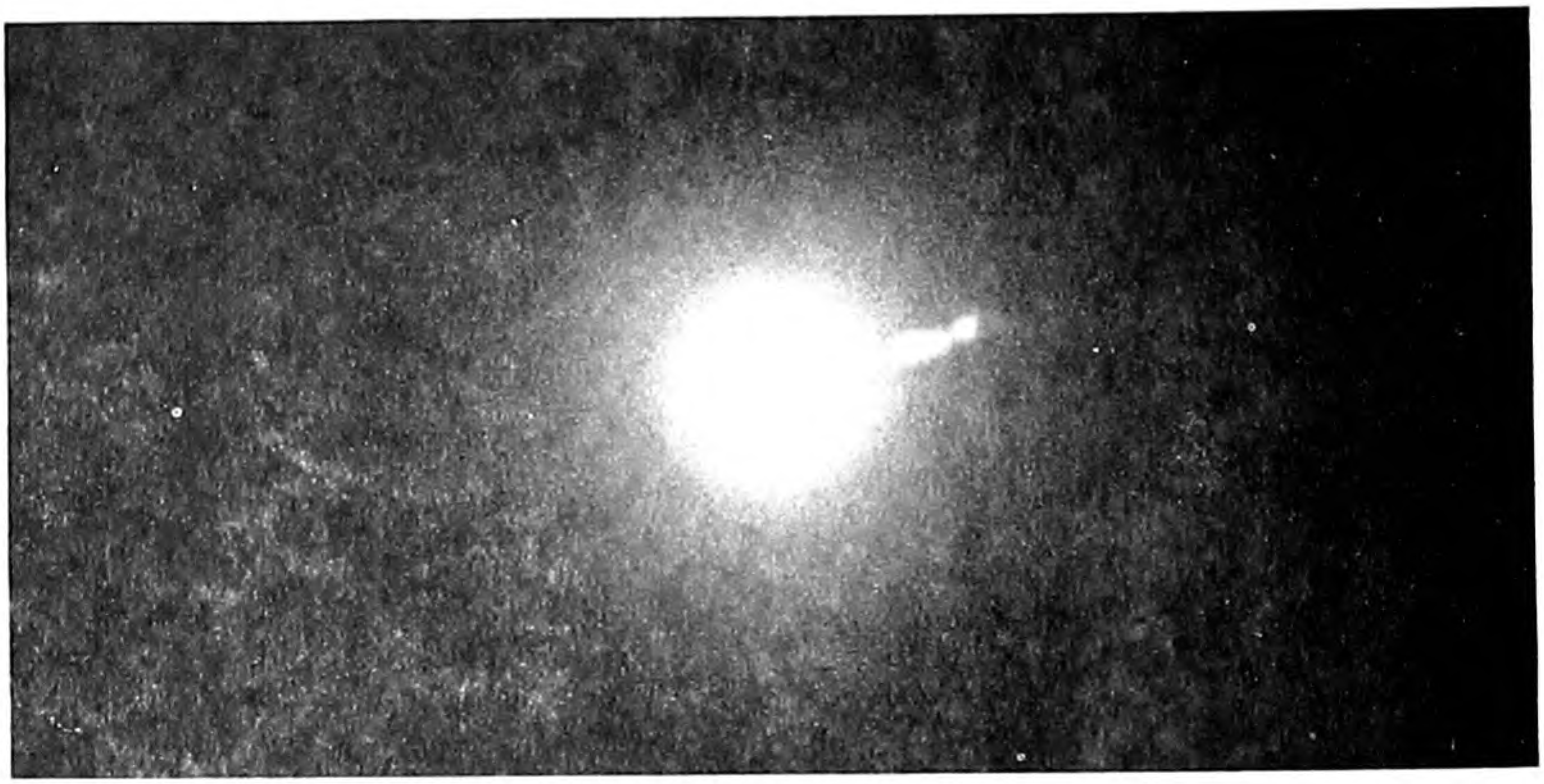
CRAB NEBULA is the remnant of a supernova which occurred in 1054. Its light comes from highly accelerated charged particles, which may have been created by the annihilation of anti-matter. This photograph is from the 100-inch telescope on Mount Wilson.

CASSIOPEIA A, the strongest isolated source of radio waves in the sky, is seen as a series of faint luminous wisps in this photograph from the 200-inch telescope on Palomar Mountain. It is probably also the remnant of a supernova and may contain anti-matter.

comes reasonable to ask whether symmetry may be preserved after all by the existence in some other part of the universe of an equal amount of anti-matter with the opposite "handedness."

We may sum up as follows. Anti-matter may exist in our galaxy, but it cannot exceed about one part in 10,000,000 of ordinary matter if it is there. It is most unlikely that any of the stars in our galaxy can be made of anti-matter. Outside our galaxy, other galaxies in remote parts of the universe may consist entirely of anti-matter. The nearest approach to direct proof of the existence of such bodies is the presence of strong radio sources whose energy is difficult to explain by any known process but might be explained by the annihilation of anti-matter. On the other hand, if anti-matter does exist in the universe, we do not understand at present how the bulk of it became separated from matter. To explain this would apparently require a revolution in our thinking on cosmological problems.

CYGNUS A, a strong emitter of radio waves, may be a pair of colliding galaxies which appear as the blob in the center of this 200-inch-telescope photograph. Its observed radio energy can be calculated if the galaxies are assumed to contain some anti-matter.



MESSIER 87 is apparently a single galaxy, but the radiation of light and radio waves by the bright jet at right is unusually strong. This radiation may be caused by the annihilation of anti-matter in the jet. The photograph was made with the 200-inch telescope

The Authors

GEOFFREY BURBIDGE and FRED HOYLE have worked together on various problems in astrophysics. Burbidge graduated from the University of Bristol and acquired a Ph.D. in meson physics from the University of London in 1951. While studying physics, he married an astronomer, which converted him to astrophysics. Burbidge has been an Agassiz Fellow at Harvard College Observatory, a researcher at the Cavendish Laboratory of the University of Cambridge, and a Carnegie Fellow at the Mount Wilson and Palomar Observatories; he now teaches at the Yerkes Observatory of the University of Chicago. Hoyle is also an astrophysicist, but his background is more mathematical than physical. By the age of six he knew the multiplication table up to 12 times 12. Failing eyesight caused him to give up cricket when he was 13, but did not prevent him from staying up all night with a three-inch telescope his parents had bought for him. Hoyle's county (Yorkshire) gave him a scholarship to the University of Cambridge, where he soon won a prize fellowship to St. John's College. Now a senior fellow there, Hoyle is widely known for his books *The Nature of the Universe* and *Frontiers of Astronomy*.

Bibliography

ANTINEUTRONS PRODUCED FROM ANTIPROTONS IN CHARGE-EXCHANGE COLLISIONS. Bruce Cork, Glen R. Lambertson, Oreste Piccioni and William A. Wenzel in *The Physical Review*, Vol. 104, No. 4, pages 1,193-1,197; November 15, 1956.

THE ANTIPROTON. Emilio Segrè and Clyde E. Wiegand in *Scientific American*, Vol. 194, No. 6, pages 37-41; June, 1956.

COLLIDING GALAXIES. Rudolph Minkowski in *Scientific American*, Vol. 195, No. 3, pages 125-134; September, 1956.

THE CRAB NEBULA. Jan H. Oort in *Scientific American*, Vol. 196, No. 3, pages 52-60; March, 1957.

THE HYPOTHESIS OF THE EXISTENCE OF CONTRATERRENE MATTER. V. Rojansky in *The Astrophysical Journal*, Vol. 91, No. 2, pages 257-260; March, 1940.

MATTER AND ANTI-MATTER. G. R. Burbidge and F. Hoyle in *Il Nuovo Cimento*, Vol. 4, No. 3, pages 558-564; September, 1956.

SPECULATIONS ON COSMOGONY. M. Goldhaber in *Science*, Vol. 124, No. 3,214, pages 218-219; August 3, 1956.

SCIENTIFIC AMERICAN November 1958, Vol. 199, No. 5, pp. 44-50 OFFPRINT 203

STELLAR POPULATIONS

by Geoffrey and Margaret Burbidge

In 1943 Walter Baade observed that stars are divided into two populations. His concept opened up a broad field of study, and now astronomers classify stars into five populations instead of two.

During the fall of 1943 there were a few nights of almost ideal observing conditions at the Mount Wilson Observatory. The air was exceptionally steady, the temperature nearly constant and the surrounding valley lay in the darkness of the wartime blackout. Walter Baade of the Mount Wilson staff seized the opportunity for another try at a long-standing problem—to photograph separate stars in the central region of the great spiral nebula in Andromeda. All previous photographs with the 100-inch telescope on Mount Wilson, the most powerful instrument then available, had shown this region only as a hazy blur of light.

The combination of favorable circumstances and of Baade's great technical skill and ingenuity was successful. His now-famous plates revealed that the blur was actually a dense mass of faint, reddish stars. He also succeeded in resolving stars in the two small companion nebulae of the Andromeda nebula: M 32 and NGC 205. As Baade himself has explained in SCIENTIFIC AMERICAN [see "The Content of Galaxies"; September, 1956], a study of these newly resolved objects suggested that all stars are sharply divided into two classes: one, which Baade called Population I, whose brightest members are hot, blue stars; and one, which he named Population II, whose brightest stars are cool and red, but very large. He concluded that the stars of Population I were relatively young, while the stars of Population II were quite old. Most of the stars near the sun, and those which had long been visible in the arms of the Andromeda spiral, fell into the first group; those in the central region of the Andromeda nebula, and in the so-called globular clusters of our own galaxy, into the second.

After 15 years of investigation our picture of stellar populations has grown less simple and perhaps less surprising. Instead of supposing that all stars fall into just two groups, widely separated in age and location in the galaxies, we now believe that there is a more nearly continuous spectrum of ages, from very ancient stars to those still in the process of birth. We still divide them into classes, but these are more numerous, and one tends to merge into the next. To appreciate how the current view has developed, let us retrace the steps since Baade's original findings.

The brightest stars in his Population I are bluish white, with surface temperature of some 30,000 degrees absolute, and the brightest of them shine with the brilliance of 100,000 suns. As we have said, the great majority of stars near the sun seemed to belong to this group, as do the stars which had been resolved on photographs of the arms of the Andromeda nebula and other nearby spiral galaxies.

On the other hand, no members of Population I were found between the arms or in the central regions of spiral galaxies, or in the so-called elliptical galaxies, which have no spiral arms. In these regions the brightest stars (in Baade's Population II) are 50 to 100 times fainter than those in Population I, and their surface temperatures are relatively low—only 3,000 or 4,000 degrees. Their color is distinctly red, and Baade was struck by their similarity to the brightest stars in the globular clusters of our own galaxy. These dense clusters, each containing some 100,000 stars, are distributed around the galaxy in a roughly spherical volume [*see illustration at top of page 21*].

Thus far we have spoken only of the brightest stars in the various regions. Each population also contains a whole array of fainter members, which seemed to fall into the same population grouping. All these stars can also be classified in another way: by means of the well-known temperature-luminosity diagram, in which intrinsic brightness is plotted against temperature. When this is done, the stars fall into a well-defined pattern, with the two populations occupying different parts of the diagram [*see illustration on page 19*]. The band running from upper left to lower right is known as the "main sequence." There are Population I stars along its entire length, but members of Population II are found only below a certain point near the middle.

Now it is known that the brightness of stars on the main sequence depends on their mass. In fact the brightness increases as the square of the mass at the lower end of the main sequence, and as the third or fourth power at the upper end. If one star has twice the mass of another, it will be four to 16 times as bright. But the mass of a star is a measure of the amount of fuel it has available to burn in the thermonuclear reactions which produce its radiant energy, and the brightness is a measure of its rate of burning. Therefore the lifetime of the star (the time required to consume all of its nuclear fuel) is proportional to the

LUMINOUS CLOUD of interstellar matter in the constellation of Scutum Sobieski was photographed with the 200-inch telescope on Palomar Mountain. The matter is made luminous by the hot young stars of Population I embedded in it. In such regions stars are probably being formed at the present time. Small dark patches of dense matter can also be seen; some of them may represent an early stage in the formation of stars.

GREAT NEBULA IN ANDROMEDA was photographed by the 48-inch Schmidt telescope on Palomar Mountain. The arms of this spiral galaxy, the disk of which is seen at an angle, are outlined by the dark lanes that lie between them. Population-I stars are found in the arms; Population-II stars, in the bright central region. The small blobs above and below the disk are small satellite galaxies.

mass divided by the brightness. Because brightness increases so much more rapidly than mass, the bright, hot stars at the upper left in the main sequence must burn themselves out much faster than the fainter stars do. In fact, the brightest and hottest appear to be less than a million years old.

Thus Baade's Population I contains relatively young stars. They are so young as a group that some must even now be in the process of formation. Indeed, we can probably see this happening in our own galaxy. There are some faint, irregularly flickering stars in the Great Nebula in Orion whose unsteadiness is almost certainly due to their youth. They have not settled down to an orderly existence on the main sequence. Some of them may even be growing yet, drawing to themselves more of the surrounding gas and dust.

We can now understand why Population I stars exist only where Baade found them, in the spiral arms of galaxies. Formed recently, they have not had time to move away from the region containing the raw materials out of which they were made. And for some reason, probably having to do with magnetic fields, interstellar gas and dust are concentrated in spiral arms. Between the arms and in the central parts of spiral galaxies, as well as in the whole of elliptical galaxies, there is no dust and very little gas.

These regions are the domain of Population II. There is good reason to suppose that they have been dust-free for a very long time, and so all the Population II stars must be quite old. It is easy to see, then, why their main-sequence members extend only to stars slightly brighter, and about 20 per cent heavier, than the sun. All the brighter ones that must have been on the upper part of the main sequence came to the end of that phase of their lives long ago, and there has been no material to make replacements.

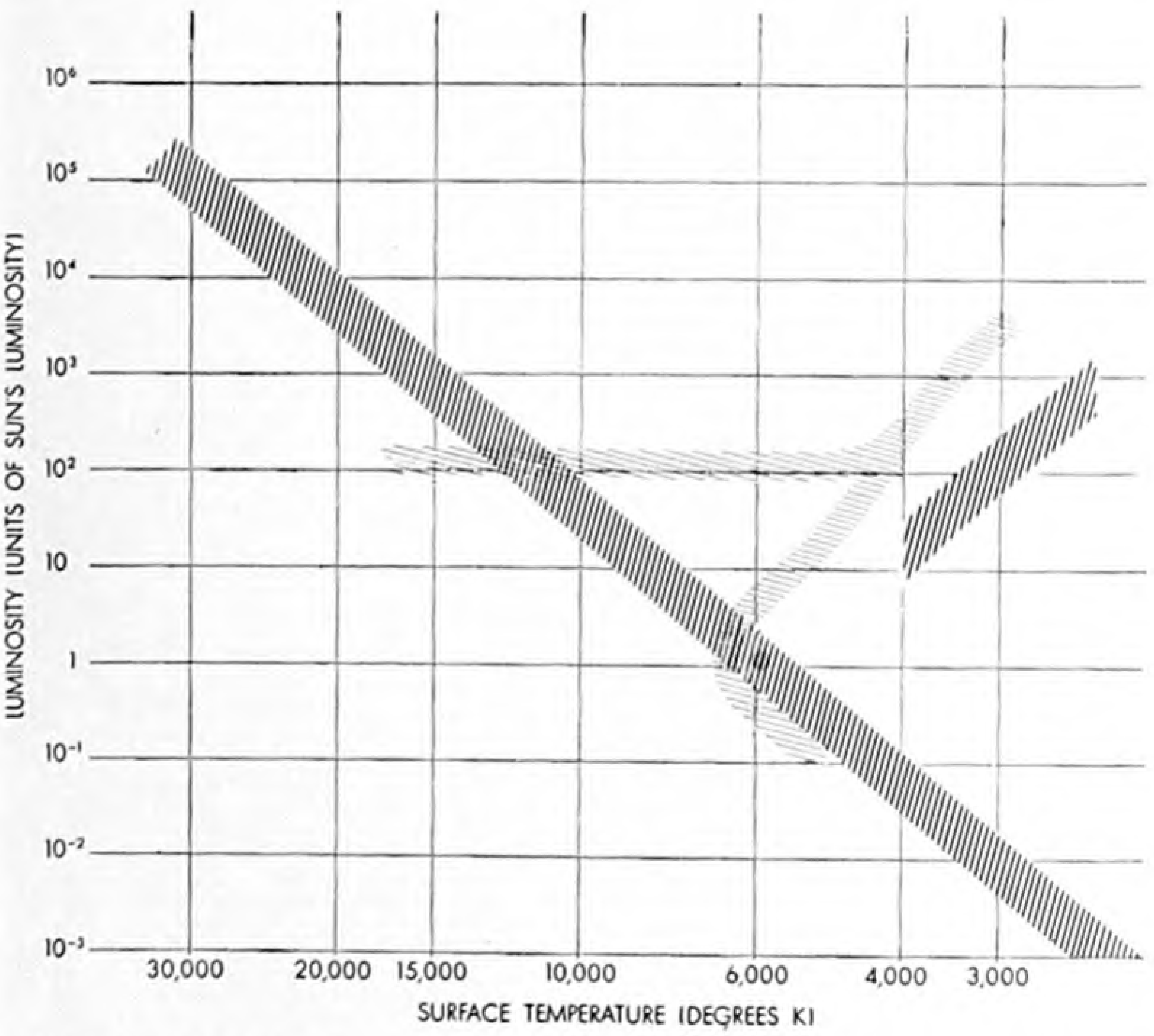
How does a star leave the main sequence? Theories of the nuclear reactions in stars show that most of their hydrogen is consumed and converted to helium deep in their interiors. As time goes on, the core in which the hydrogen has been totally consumed grows larger and larger. When the core comes to contain about a 10th of the whole mass of a star, the star's internal structure becomes unstable. To restore equilibrium its material must be rearranged. In the process the star expands fairly quickly, and its surface layers cool; it becomes a "red giant." This is what has happened to the red stars that are the brightest members of the globular clusters, and to those which appear on Baade's photographs of the Andromeda nebula and of its companions.

Eventually a red giant uses up all the nuclear fuel in its center and comes to the end of its life as a normal star. Then it may become a supernova, disintegrating in a giant thermonuclear explosion, or it may suffer a series of lesser explosions known as nova outbursts. Such must have been the fate of the bright stars of Population II. In Population I there are also some red giants, but they are about 15 times less bright than those of Population II. Examination of their spectra has shown that the chemical compositions of the two types of red giants are not the same. In the Population II red giants the heavier elements such as calcium and iron are only about one one-hundredth as abundant with respect to hydrogen as they are in the red giants of Population I. Thus the members of Population II must have been made out of material that was relatively poor in the heavier elements.

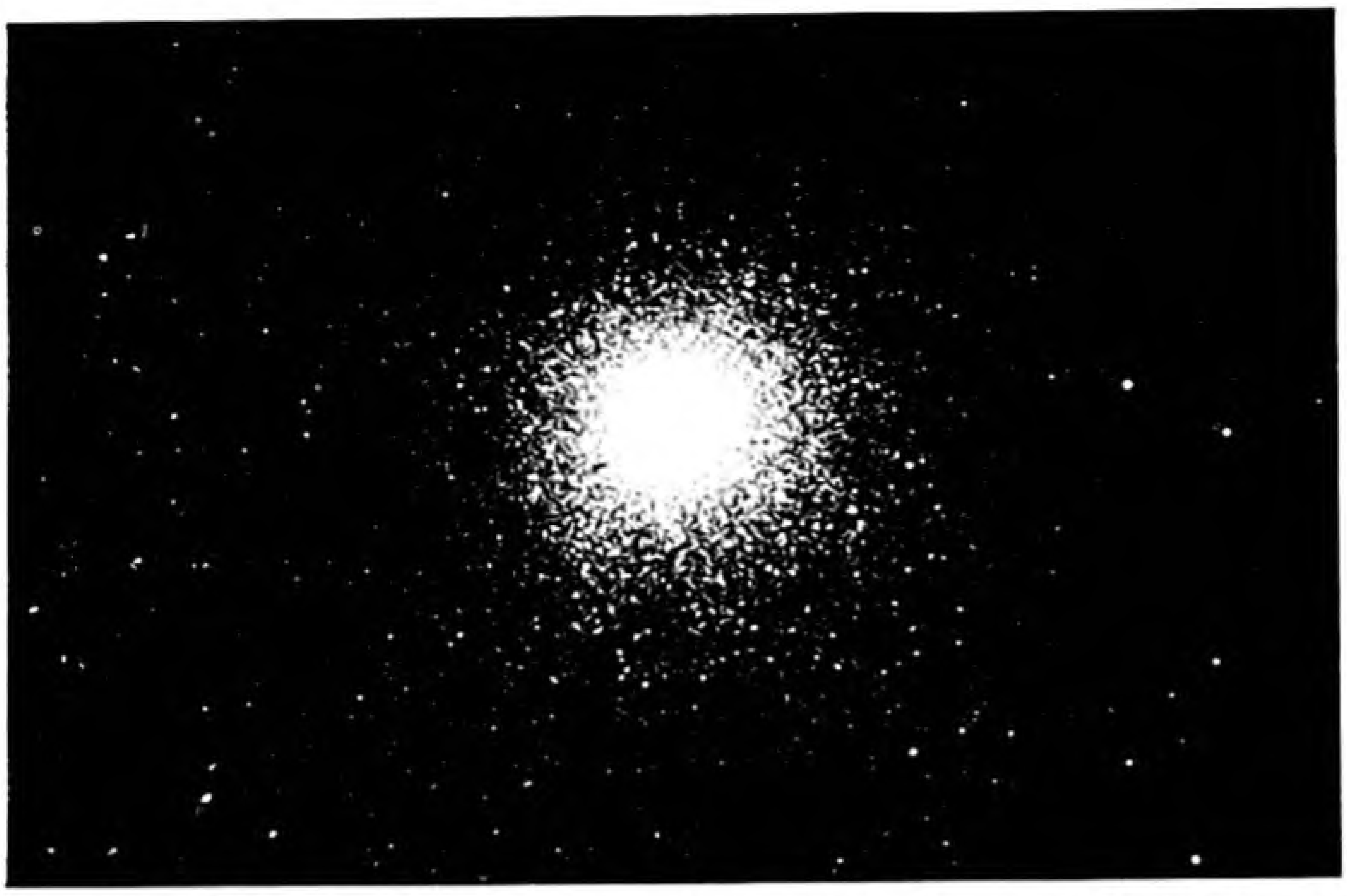
This is just what we should expect from the current theory of the origin of the elements. It tells us that in the beginning there was only hydrogen. All the other chemical elements have been created out of hydrogen by nuclear reactions in stars. Each time a star goes through its explosive death throes it spews out the heavy elements it has manufactured during its life. Hence the dust and gas out of which new stars are made must have been gradually enriched in the heavier elements. Thus it is not surprising that the oldest stars we observe today should contain the lowest proportion of these substances.

All this seems reasonable enough, but it does not explain why there should be only two stellar populations, an old one and a young. Why not some middle-aged stars? As we have indicated, the sequel to Baade's work has resolved the puzzle. Middle-aged stars do indeed exist.

The most revealing indication came from studies of the movements of stars in our own galaxy. The galaxy as a whole is revolving, and the individual stars share in this motion. In addition they have movements of their own. Not only do they travel about in the central plane of the galaxy, but most of them have a component of motion perpendicular to the plane [*see illustration at the bottom of page 21*]. This component



TEMPERATURE-LUMINOSITY DIAGRAM elucidates the relationship between stars of Population I and those of Population II. The black hatched areas are occupied by stars of Population I; the colored hatched areas, by stars of Population II. The red-giant stars are at the upper right. The sun, a star of Population I, is indicated by the colored dot.



GLOBULAR CLUSTER, M 13 in the constellation of Hercules, was photographed with the 200-inch telescope. A member of our own galaxy, it consists only of Population II stars.

ELLIPTICAL GALAXY, NGC 205, photographed with the 200-inch, is composed of Disk-Population or Population-II stars. It is the object near the top of photograph on page 18.

(which, in the reference system conventionally adopted in astronomy, is along the z axis) carries them above or below the central plane. They can travel only so far in the z direction before the gravitational pull of the central mass of stars brings them back. Then they reverse their motion and move back toward the plane. Like pendulums, they do not stop at the equilibrium point, but overshoot and travel an equal distance to the other side of the plane, and so on.

At any given moment the z-speed of a star depends on its position in the oscillatory cycle. But, in general, stars with the highest z-speed should swing farthest from the central plane. And so it turns out. The globular clusters, arranged nearly spherically around the plane, have z-speeds of about 100 kilometers per second. The brightest, hottest stars, which lie in a very flat disk in the central plane have z-speeds of only about five kilometers per second. But the catalog of speeds is not restricted to these extreme values. There is a continuous gradation; and in the case of main-sequence stars the brighter they are, the smaller their speeds and the nearer they lie to the central plane. This fact emerged gradually, in the course of many years of observation by several astronomers. In 1950 the Soviet astronomer P. P. Parenago drew attention to it and suggested that, rather than two clearly defined populations of stars, there must be a full range of populations.

More recently we have found that the chemical compositions of stars show a similar spread. The percentage of heavy elements varies from the very low value characteristic of the globular clusters to the much higher value found in the brightest stars. Thus our galaxy appears to contain stars of various ages, with the older ones lying, on the average, farther from the central plane.

The picture is quite satisfactory because it fits well with the current view of galactic evolution. We suppose that our galaxy began its life as a cloud of hydrogen gas, either pure or slightly contaminated with heavier elements from exploded stars in earlier galaxies. The cloud then began to shrink, pulled together by its own gravitational attraction. As it did so, the first stars or clusters of stars began to form. In the beginning the gigantic cloud probably revolved slowly, speeding up as it shrank. As a consequence of its rotation the sphere gradually flattened, under the same kind of force that makes the rotating earth slightly flattened at its poles.

All the time the cloud was shrinking,

flattening and revolving faster and faster, stars were forming in those regions where, by chance, the density was higher than average. Big stars went quickly through their life histories, cooking up heavy elements and then exploding and scattering them back into the gas out of which new stars were continually condensing. The shape of the cloud at any epoch should be preserved by the stars formed at that time. Once they had coalesced into dense masses, they would move more or less independently of the surrounding gas and dust, and would no longer partake in the general flattening and shrinking. As we have seen, the oldest stars we see today are indeed distributed most nearly spherically.

At a conference in Rome in the summer of 1957 astronomers generally agreed on a convenient classification of stars into five populations. They are as follows:

1. Extreme Population II. This is the oldest group—at least seven or eight billion years old. In our galaxy it is represented by the globular clusters, together with a sparse spherical distribution of stars lying between them. These isolated stars may have escaped from the clusters.

2. Intermediate Population II. Not quite as old as the first group, it occupies a volume not completely spherical, though not flattened very much. Stars which explode as novae, and also the so-called planetary nebulae, apparently belong either to this or to the next group.

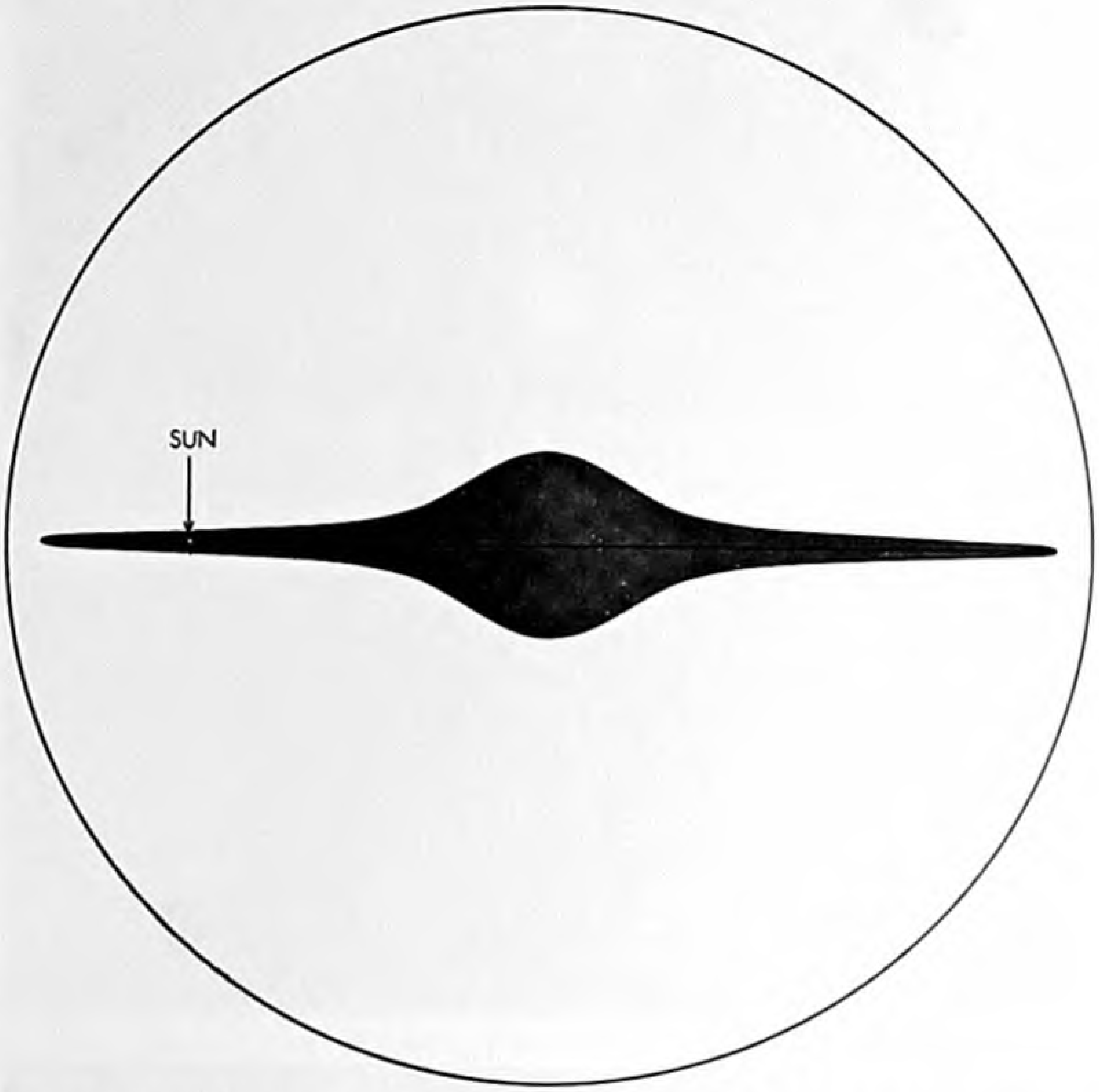
3. Disk Population. These stars probably range from three to five billion years old. The Disk Population makes up the great bulk of the stars in our galaxy and in the Andromeda nebula; most of the stars between the spiral arms and in the dense central regions probably belong to it. The sun is probably a member. We know from its chemical composition the sun must be at least a "third generation" star, which indicates that many stars completed their life cycles before the Disk Population was formed.

4. Intermediate Population I. This Population ranges from about a hundred million to a few billion years old. Its members, which include stars like Sirius, lie in or quite near the central plane of the galaxy but are not restricted to spiral arms.

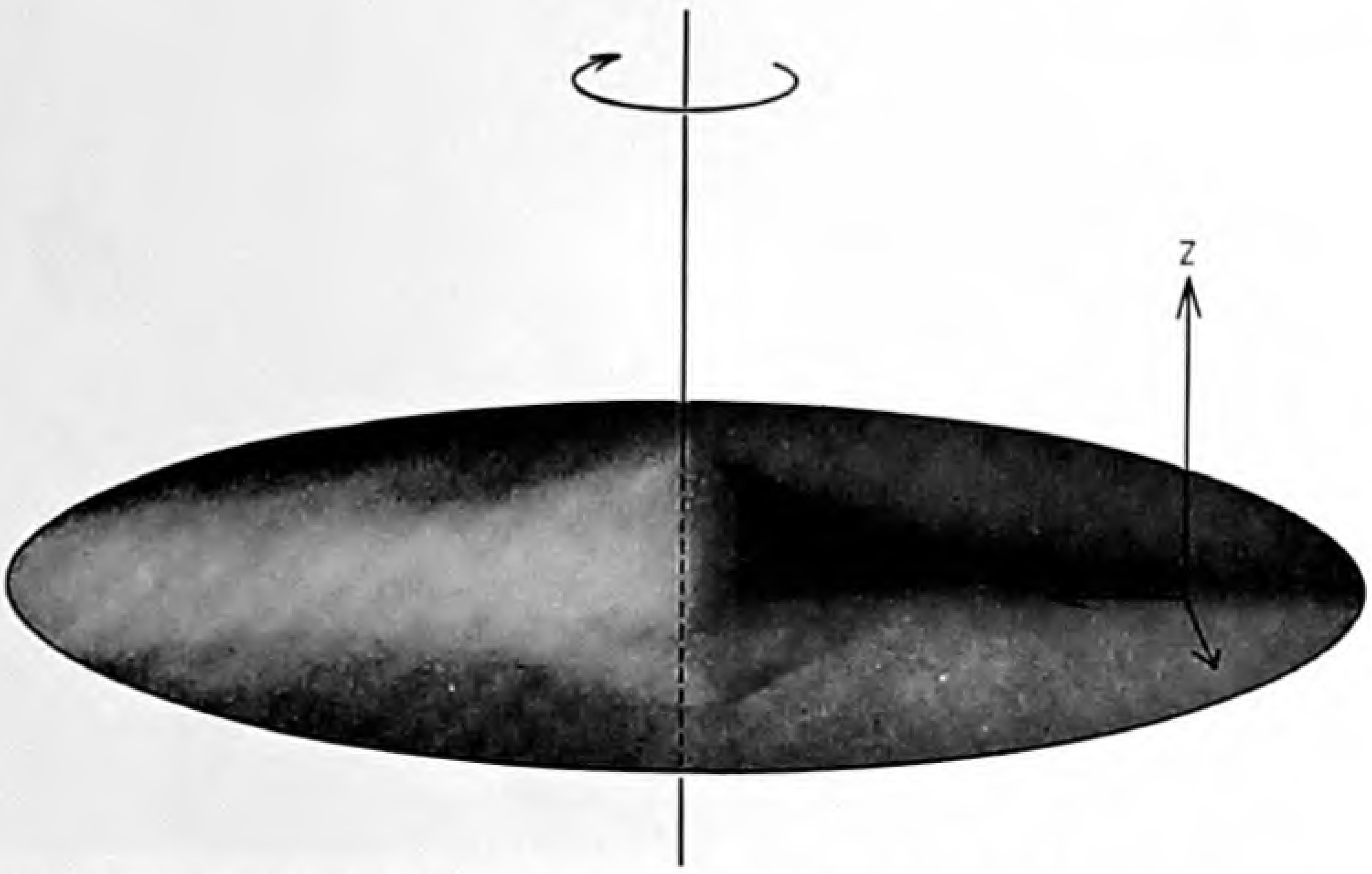
5. Extreme Population I. This is tens of millions of years old or less, and it includes the gas and dust still not condensed into stars. Both the gas and the stars lie in spiral arms. The hot, bright stars in Orion, particularly Rigel and the stars in the Orion nebula, belong to this class.

This division into five populations is a matter of convenience; actually the groups merge into one another and could be further subdivided.

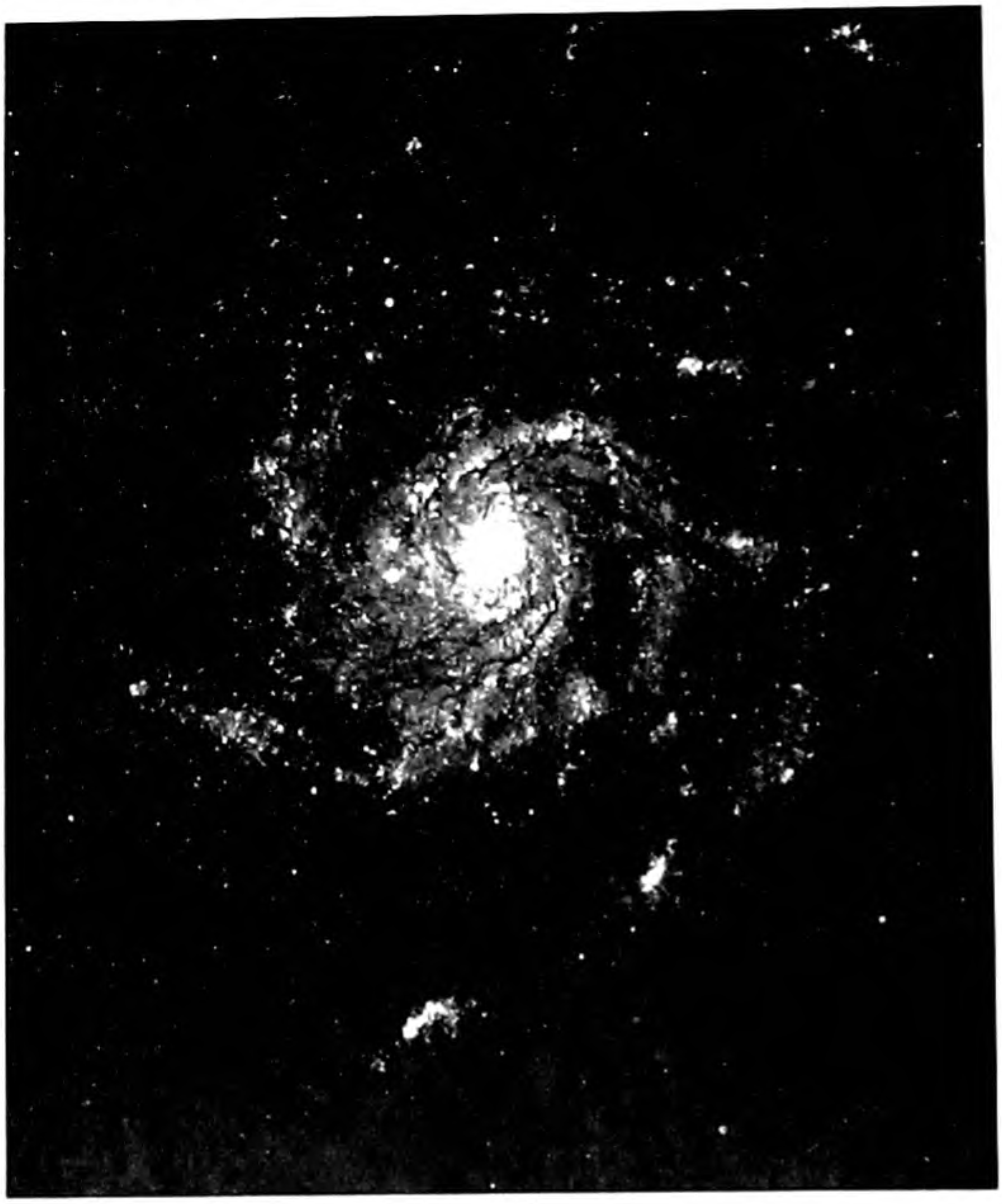
According to the new grouping, stars in the central regions of the Andromeda nebula and of the large elliptical nebulae are Disk Population, while the globular-cluster stars belong to Extreme Population II. The chemical compositions seem to support the classification: globular-cluster stars are considerably

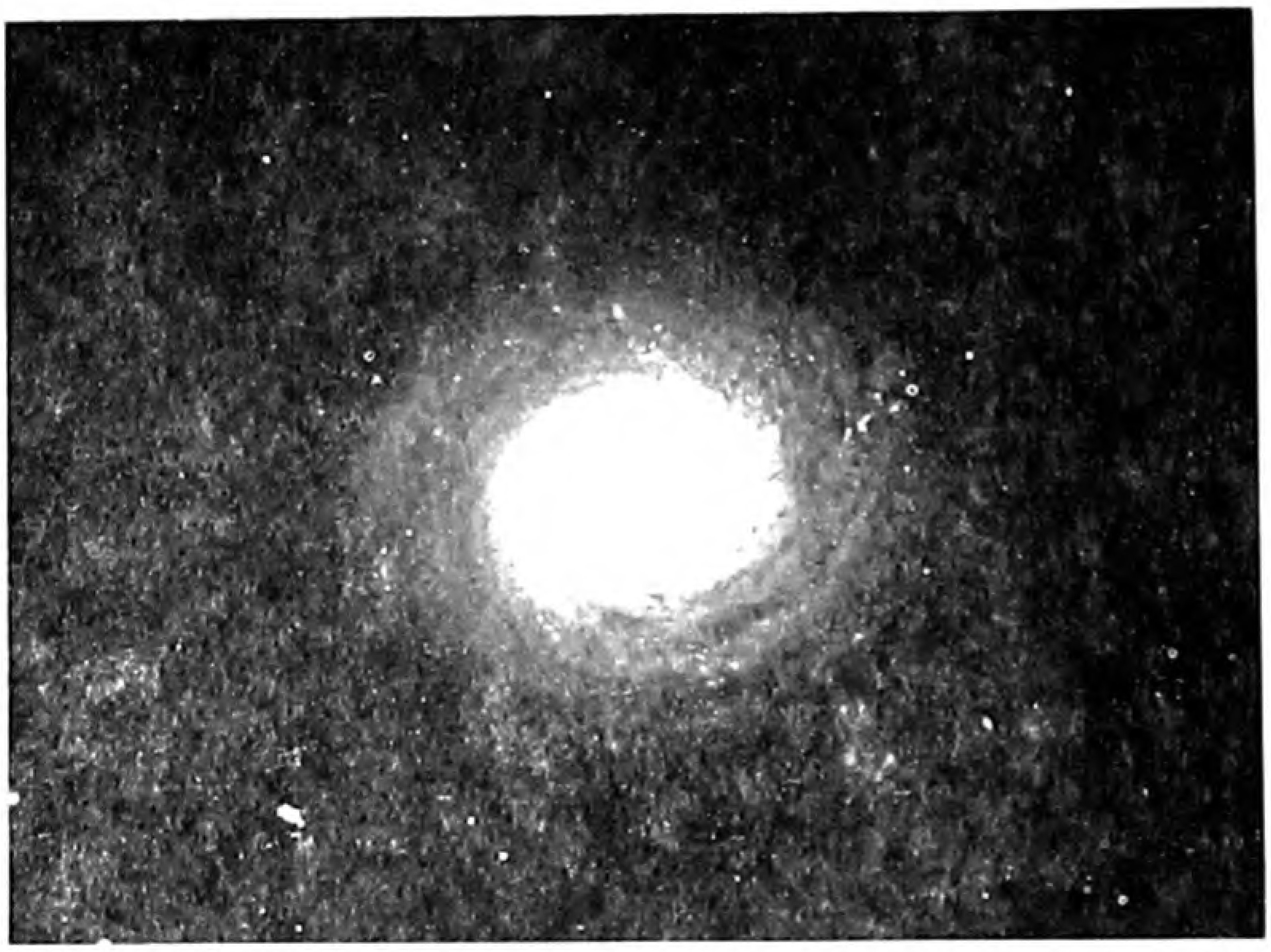


SCHEMATIC EDGE-ON VIEW of our own galaxy shows its central bulge and relatively thin arms. The circle is a cross section of the spherical volume occupied by the globular clusters.



SCHEMATIC THREE-QUARTERS VIEW of the galaxy shows the three components in the motion of a star in its central plane. The star swings like a pendulum in the z direction, while it also moves in the xy plane, taking part in the rotation of the galaxy as a whole.





CENTRAL REGIONS of two spiral galaxies are compared. At top is the galaxy NGC 5457; its central region is small in comparison to the extent of its arms. At bottom is NGC 4736; its central region is large in comparison to the extent of its arms. Since the central regions of galaxies contain old Disk-Population or Population-II stars, the second galaxy may be older than the first. Both galaxies were photographed with the 200-inch telescope.

poorer in the heavy elements than the others. But there are still some open questions. We have mentioned that, because of their lower concentration of heavy elements, the bright red giants in globular clusters are about 15 times brighter than those near the sun. Yet the brightest red giants in the center of the Andromeda nebula apparently have the same brightness as those in globular clusters. Also, in the central regions of our galaxy there is a large number of variable stars of the same type as those found in globular clusters. Perhaps these regions actually contain a mixture of classes, including Extreme Population II, as well as Intermediate Population II and Disk Population.

Clearly we have only begun to understand the whole problem. When we learn how to classify all the stars we can observe, we shall know a great deal more about the history of the universe. Many intriguing questions suggest themselves. For example, what is the detailed life history of a galaxy? Are the galaxies around us in different stages of development? It seems now that they are. Spiral galaxies with very small central bulges may be much younger than galaxies like ours, in which the central bulges are large. Probably the central region grows larger as a galaxy ages, because its material gradually loses its random movement and falls inward. As the center becomes denser, star formation speeds up. Therefore it is not surprising to find stars covering a considerable range of ages in the central regions.

It is possible that all galaxies do not age at the same rate. Factors such as the mass of a cloud of gas, its initial speed of rotation and the size of its magnetic field, if it had one, may affect its development and the rate of star formation in it. Thus galaxies that were born at the same time may now have reached very different stages in their life histories. This might explain why old- and young-looking galaxies are sometimes found very close together. For example the two Clouds of Magellan, our nearest extragalactic neighbors, seem quite young as compared with our galaxy. Half of their masses are still in the form of gas, whereas the gas in our galaxy comprises only a few per cent of its mass. Whether they are also poorer in heavy elements is not yet certain.

Many different branches of astronomical research are at present converging on the problem of the life histories of galaxies. But the concept of stellar populations, originated by Walter Baade, remains the key to all the approaches.

The Authors

MARGARET and GEOFFREY BURBIDGE are a husband-and-wife team—she being an astronomer, he an astrophysicist. The Burbidges, who are English, met 10 years ago in a University of London lecture hall. At that time she was employed in the University of London Observatory, while he was a physicist. Their courtship resulted in his conversion to astrophysics and also in their marriage. In 1951, when Geoffrey Burbidge came to the Harvard College Observatory as an Agassiz Fellow, his wife became a fellow of the University of Chicago's Yerkes Observatory at Williams Bay, Wis. (the Burbidges commuted between the two places). In 1953 the Burbidges returned to England, where he worked for two years in the Cavendish Laboratory of the University of Cambridge. Since last year they have collaborated at the Yerkes Observatory.

Bibliography

FRONTIERS OF ASTRONOMY. Fred Hoyle. William Heinemann, Ltd., 1955.

THE RESOLUTION OF MESSIER 32, NGC 205, AND THE CENTRAL REGION OF THE ANDROMEDA NEBULA. W. Baade in *Astrophysical Journal*, Vol. 100, No. 2, pages 137-146; September, 1944.

STRUCTURE AND EVOLUTION OF THE STARS. Martin Schwarzschild. Princeton University Press, 1958.

SCIENTIFIC AMERICAN July 1955, Vol. 193, No. 1, pp. 80-87 OFFPRINT 204

DISLOCATIONS IN METALS

by Frank B. Cuff, Jr. and L. McD. Schetky

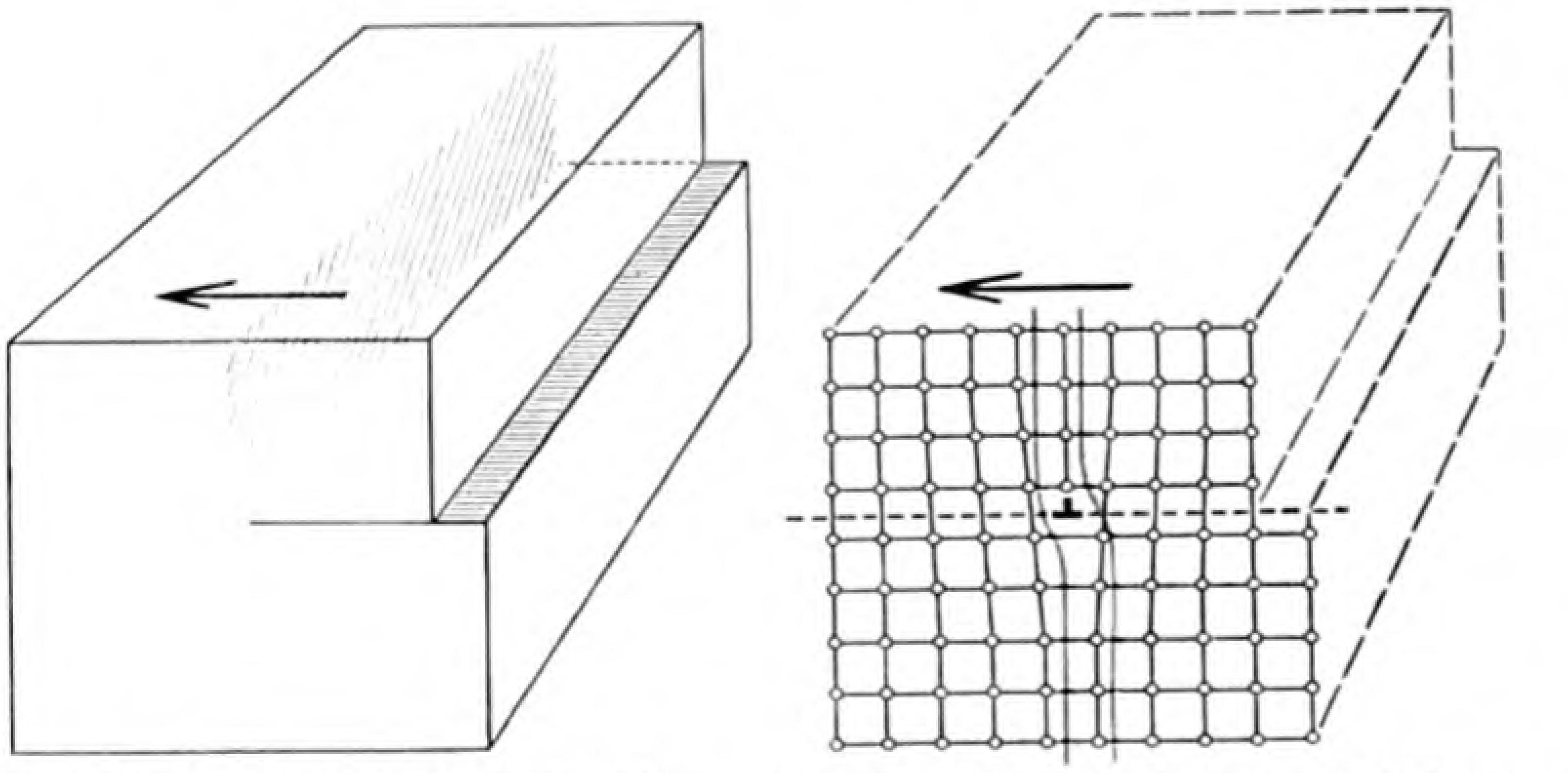
The strength of metals, and other important aspects of their behavior cannot be explained if they are considered to consist of perfect crystals. Dislocations are crystal faults that account for many of these properties.

The word "crystal" conjures up a picture of infinite orderliness—billions upon billions of identical atoms stacked in perfect array. Actually, of course, nothing is perfect. In a real metal or other crystalline substance the order is marred by a missing atom here, a foreign particle there and other departures from regularity. However, these imperfections are rare, and one might suppose that they could be disregarded in calculating the properties of the material.

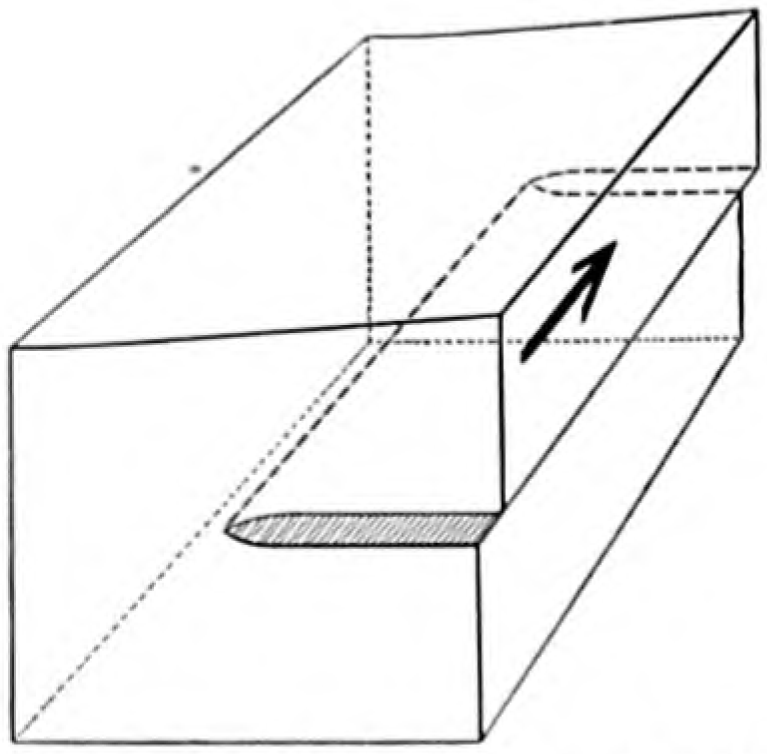
For a long time they were. On the assumption of a perfectly filled crystal lattice, metallurgists were able to make a number of accurate predictions about metallic behavior. They also got answers that were entirely wrong. One of the most annoying failures came in trying to account for the strength—or rather the weakness—of metals. For example, the theory said that a stress of some two million pounds per square inch would be required to deform pure iron plastically, that is, beyond the limit of elastic recovery. In fact it takes only 30,000 pounds per square inch.

This discrepancy, and others even bigger, remained wholly unresolved until about 20 years ago. Then it was suggested that resistance to plastic deformation depends not on the average properties of the almost perfect lattice, but on the individual properties of a previously unknown kind of imperfection. This weakest link, known as a dislocation, was proposed independently in 1934 by G. I. Taylor in England and E. Orowan, then in Germany and now at the Massachusetts Institute of Technology. Since then the hypothesis has become the basis for an entirely new and promising theory of the plastic behavior of metals. A large group of properties which previously could be studied only in an empirical way are now beginning to yield to rational analysis.

The dislocation idea is most readily visualized in terms of the deformation it was invented to explain. The simplest type of plastic deformation in a crystal may be likened to the distortion produced when a perfectly stacked deck of cards is pushed askew. The planes of atoms, like the cards, slip over one another. In the old theory the stress necessary to produce this deformation was calculated as the force required to slide whole planes of atoms over each other. Now suppose that the planes do not actually move as rigid units. Imagine that a slip can occur in part of a plane, as shown in the first diagram below. The next diagram shows what would happen at the atomic level. Here the upper right portion of a crystal has been moved one atom spacing to the left with respect to the lower right portion, while the left half of the crystal has remained undisturbed. Except near the boundary of the slip, the lattice arrangement is unimpaired, and the atoms are in register. In the boundary region, however, the arrangement must be upset—dislocated. Obviously there must be one more vertical plane of atoms above the slip surface than below. The imperfection which oc-



EDGE DISLOCATION (*symbolized by inverted T*) occurs at the boundary of slip between layers of a crystal, the slip direction (*arrows*) being perpendicular to the boundary line. Colored vertical plane (*left*) represents the extra row of atoms appearing above the dislocation line when the vertical layers are subject to distortion shown in the second drawing.



SCREW DISLOCATION occurs when the slip direction (*arrow*) is parallel to the slip boundary. Like the edge dislocation, it is a region of distortion and high energy.

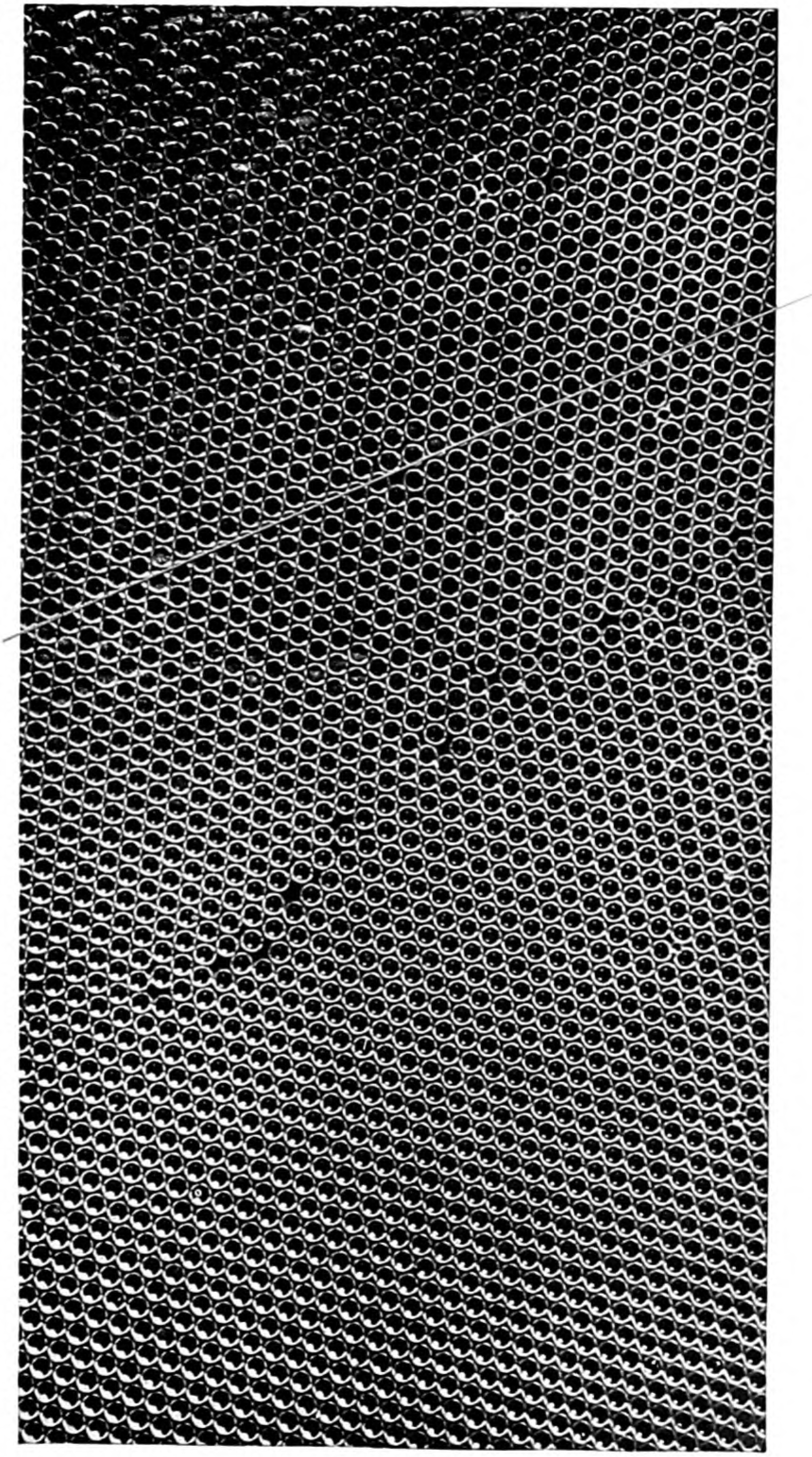
curs along the bottom of this extra row of atoms is called an edge dislocation. A model of this kind of atomic derangement is provided by the layer of bubbles in the photograph at the right.

The bubbles show what could happen in an array of atoms, but they do not prove that it does happen. For some years after Taylor and Orowan published their conjecture a number of authorities doubted the existence of dislocations. Now, as we shall see, the weight of favorable evidence has become so great as to leave no room for uncertainty. Dislocations exist, and they play a central role in determining many properties of metals. To appreciate their effect on metallic behavior we need to know a bit more about the properties of the imperfections themselves.

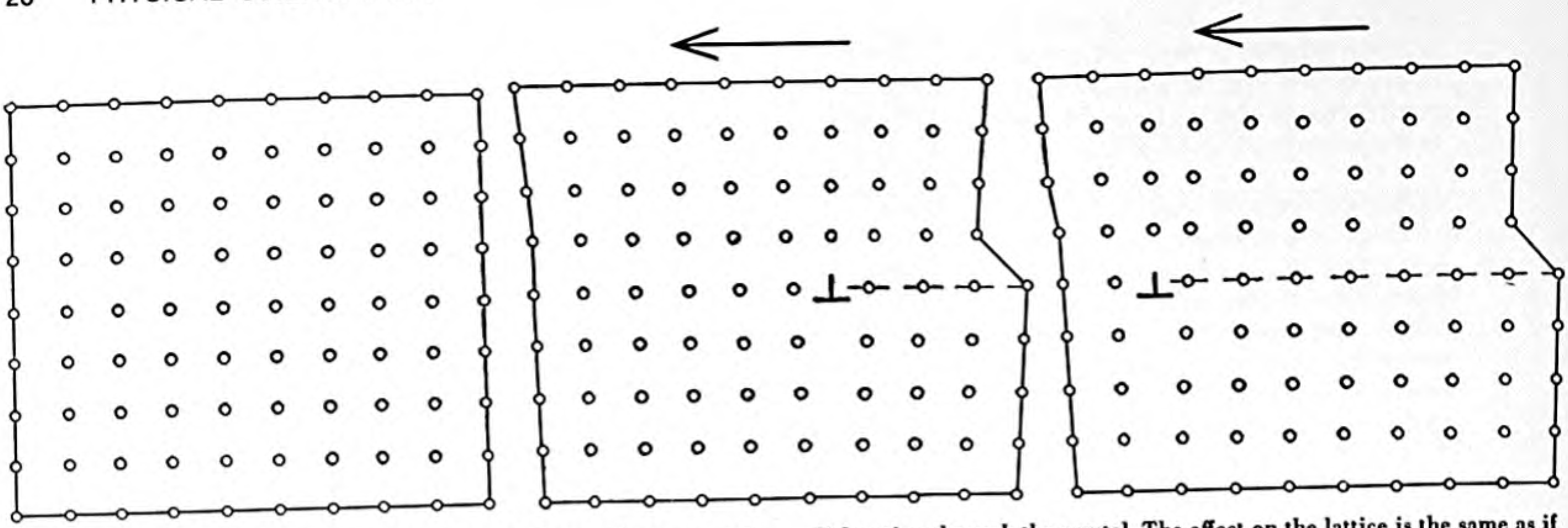
We think of a dislocation as a line running through a crystal (although it is really a region of small but finite cross section). Around such a line is a region of energy higher than in the rest of the crystal. This is because the lattice is crowded, or compressed, in the neighborhood of the extra atom plane, and pulled out, or in tension, on the opposite side. Both conditions represent an increase in potential energy over the undistorted region of the lattice. Because of nature's universal preference for the lowest possible energy states, a dislocation line acts like a stretched elastic string. It tends to be as short as possible since this makes the high-energy region as small as possible. Thus a dislocation line resists being bent or curved.

The line of an edge dislocation is perpendicular to the direction of slip. As the diagram at the right on the opposite page shows, this is not the only possibility. The slip may be parallel to the boundary of the slipped area (the dislocation line), in which case a screw dislocation is produced. This type of dislocation was discussed in detail by R. L. Fullman in his article on crystal growth [see "The Growth of Crystals," by Robert L. Fullman; SCIENTIFIC AMERICAN Offprint 260]. For our present purposes it is sufficient to note that screw dislocations are also regions of distortion and hence of high energy, and that their effects on crystal properties are the same as for edge dislocations.

What are these effects? First of all, as Taylor and Orowan originally pointed out, dislocations make metals weak. It is not hard to see why this should be so. As the diagrams at the top of the next two pages indicate, pushing one dislocation entirely across a slip plane has the effect of shifting the adjacent layers



SOAP BUBBLES floating on liquid provide a two-dimensional model of the arrangement of atoms in a crystal. The slanting colored line runs through an extra row of bubbles in the left-hand half of the picture. At the end of this row is an edge dislocation. An inch or two below the line and roughly parallel with it can be seen the bubble counterpart of a grain boundary. The bubble method was invented by W. L. Bragg and J. F. Nye. This example of the technique was prepared by Marsbed Hablanian, Massachusetts Institute of Technology.



SHEARING ACTION, in which a pair of crystal layers slip over each other by the amount of one atom spacing, is the result of moving a single dislocation through the crystal. The effect on the lattice is the same as if one whole horizontal plane of atoms had moved over the other. After

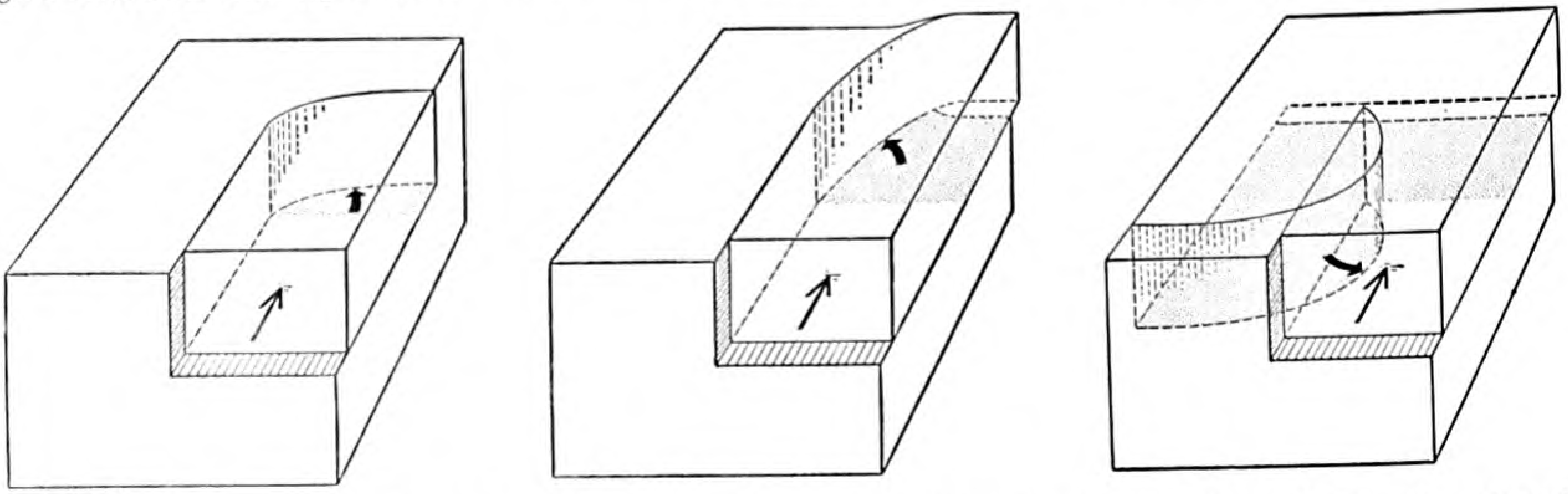
by one atom spacing. But at any moment the only atoms actually in motion are those in the region of the dislocation itself. Obviously it should be much easier to move these few particles than to push one whole plane of atoms across the other. Moving entire planes would be like dragging one sheet of corrugated iron across another; each row of atoms would have to climb a hump to drop into its new position. In a dislocation the climbing is restricted to a few rows at a time. As a matter of fact, a detailed calculation of the force necessary to move a dislocation showed that if it were the controlling factor in plastic behavior, metals would be a great deal weaker than they are. The dislocation hypothesis seemed to have done its work too well.

The explanation of this new problem is that crystals contain many dislocations which interact with each other. For example, consider what would happen if one dislocation were pushed toward another whose extra plane of atoms lay on the same side of the slip plane. (Such a pair of dislocations are said to have the same sign.) The two overcrowded regions would move closer together, thus aggravating their compression, and the stretched regions would similarly be under greater tension. Obviously there would be resistance to such a motion. Dislocations of like sign repel each other. On the other hand, if the extra atoms lie on opposite sides of the slip plane (dislocations of unlike sign) the overcrowded region of one dislocation "fits into" the stretched region of the other. These dislocations attract each other. Therefore, regardless of sign, interacting dislocations tend to immobilize themselves.

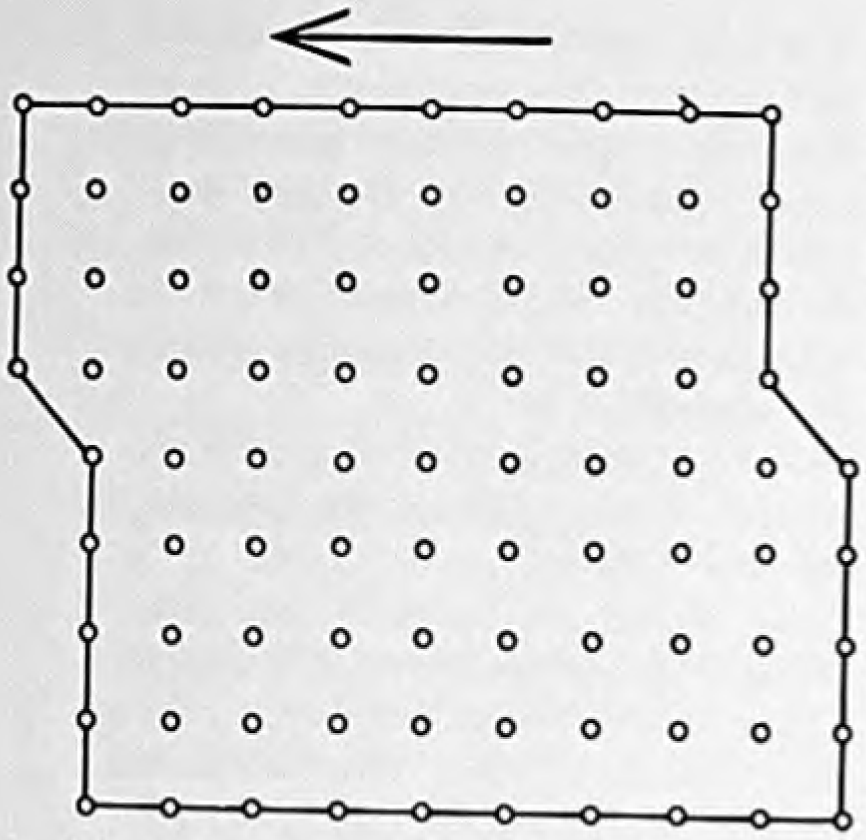
Interacting dislocations provide the first rational explanation for the effect of work hardening, *i.e.*, strengthening a metal by subjecting it to some form of plastic deformation. Imagine a series of dislocations, all of the same sign and all lying on the same slip plane, moving in response to an applied force. If the leading dislocation encounters a disturbance such as a flaw in the lattice or a foreign atom, it may be unable to move past this barrier. If the dislocation is stopped, its repulsive force will block the dislocation behind it. The dislocations will pile up like automobiles at a red light. Now no further slipping can occur unless the force is made large enough to dislodge the first dislocation. In other words, the metal is stronger.

Strengthening also results when two dislocations of opposite sign come together on the same plane. The two cancel each other, their extra atom planes combining to form a regular layer in the crystal lattice. When dislocations disappear, the path of easy deformation is lost and the metal is strengthened. Work hardening can more than treble the strength of a material. It is the basis of such standard metallurgical processes as rolling, swaging, forging and drawing.

We have seen that in the process of deformation, dislocations move to the edge of a crystal and disappear. It might be supposed that eventually all the dislocations in a metal would be lost, and that it would acquire the strength of a perfect crystal. This never happens. Apparently there is some continuing source of dislocations in every crystal. The nature of the source was a great puzzle until F. C. Frank of the University of Bristol and W. T. Read of Bell Telephone Laboratories proposed the ingenious



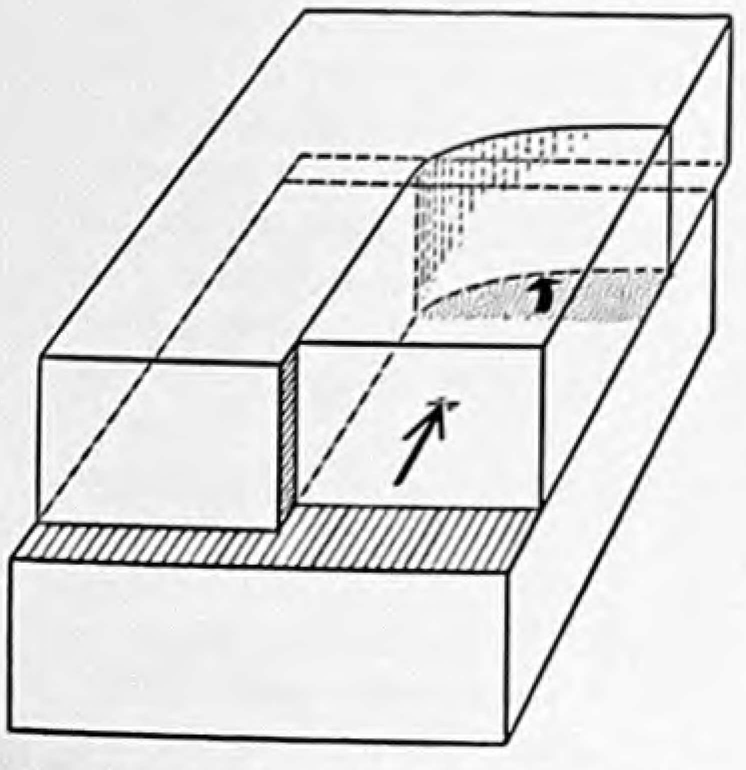
DISLOCATION SOURCE which can give rise to unlimited amount of slip is diagrammed above. The light arrow represents the direction of the applied force; the heavy arrow, the direction of motion of the horizontal dislocation line. The left end of the line has encountered a barrier



the deformation the rows of atoms on opposite sides of the slip plane are again in register (*last diagram*).

mechanism illustrated below. If a moving dislocation line meets an obstruction that stops it at an end point, the rest of the line may continue to move by pivoting around the fixed point. The diagrams show how such a line will form a spiral, which sweeps repeatedly over the slip plane, producing a slip of one atom spacing for each revolution. Thus the dislocation is never "used up." A similar sequence results if both ends of the dislocation are anchored, except that the successive waves of dislocations have the form of closed loops rather than spirals.

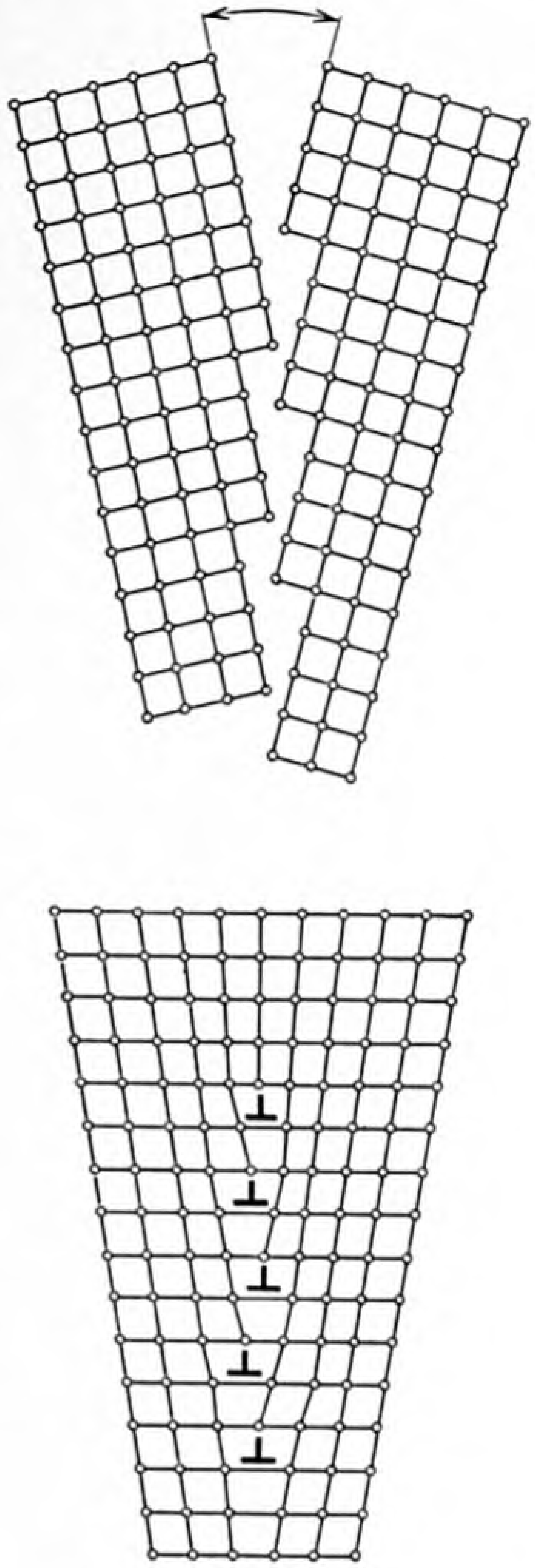
Because of their distortion of the lattice, dislocations tend to attract the foreign atoms in a crystal, such as the carbon atoms in steel. A foreign atom that is larger than the lattice atoms will tend to move to the tension side of a dislocation, where there is more room between neighbors. Similarly a smaller foreign atom will tend to migrate to the compression side of a dislocation. This effect was first pointed out by the British metallurgist A. H. Cottrell; the concentration of foreign atoms is called a Cottrell atmosphere. It explains for the first time



which arrests its motion. The dislocation pivots around this barrier point in a never-ending spiral.

the phenomenon in metals known as the yield point. When a distorting force is applied to a metal, the deformation grows steadily greater as the force increases until the yield point is reached. Then the metal suddenly gives way, and the deformation continues to increase even if the force is reduced. The yield point marks the transition from elastic to plastic behavior. The theory is that in the elastic range the forces are not large enough to pull the dislocations loose from the Cottrell atmospheres. At a certain critical value the dislocations are torn from their anchor, and may then be kept in motion by a smaller force. Metals without abrupt yield points are now being made simply by purifying them enough to eliminate the Cottrell atmosphere from all but a few of their dislocations.

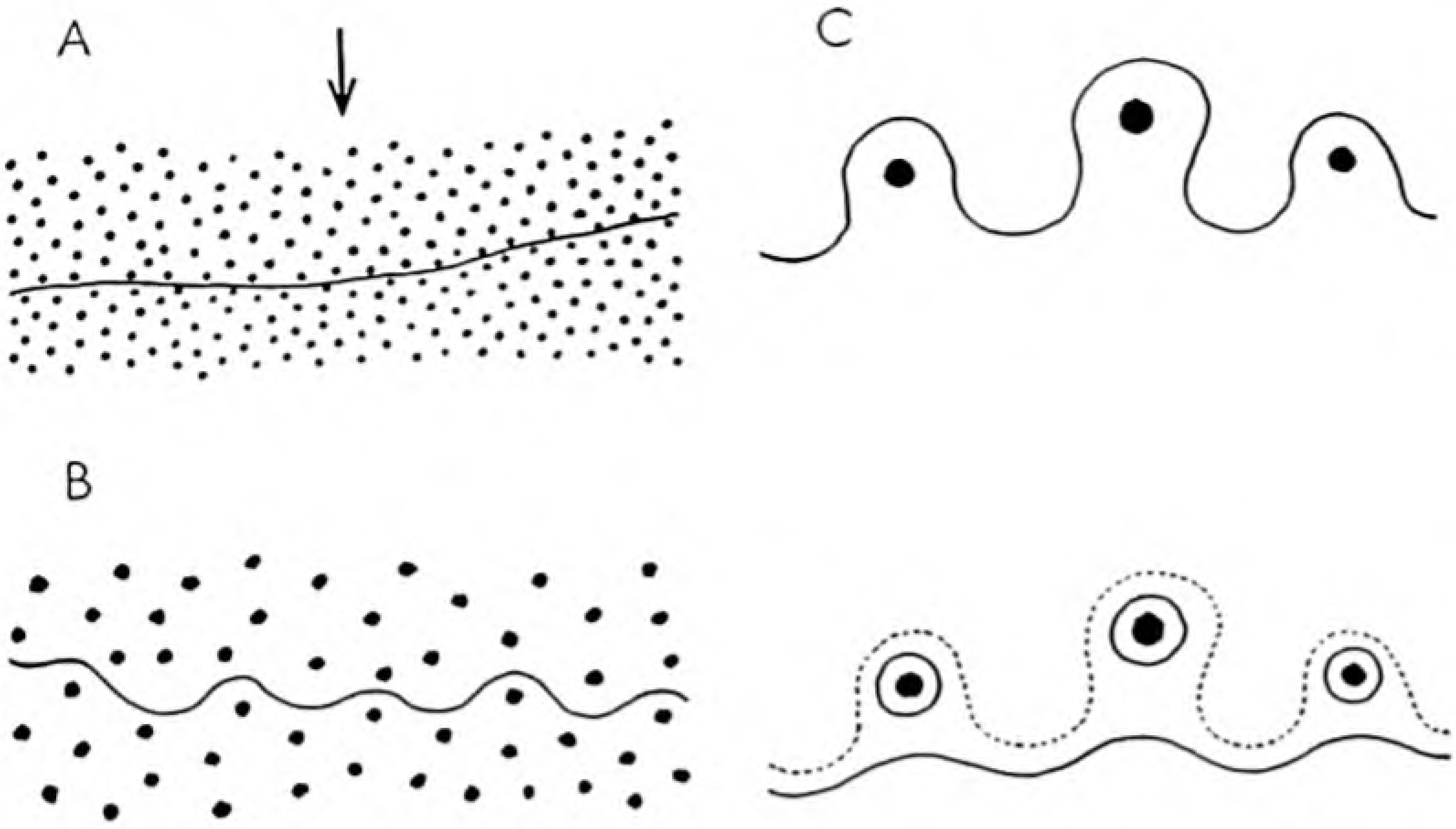
An important method for strengthening certain alloys, especially light ones such as Duralumin, is known as age hardening or precipitation hardening. In this process the metal is made very hot and suddenly quenched. Then it is held at a moderate temperature for an extended period. The result is that small particles of a second structure precipitate out of the lattice of the parent metal. Why this should strengthen the material was not clear until the advent of dislocation theory. Then it was pointed out that the precipitated particles do not fit exactly into the lattice, and so produce a region of stress around themselves. When a moving dislocation encounters such an assemblage of particles it finds that some of the stress fields oppose its passage while others tend to aid it. Before aging, the particles are very fine and closely spaced. Hence along every small section of its length the dislocation encounters about as many helping stresses as hindering stresses [*see diagram at top left on page 28*]. The effect of aging is to consolidate the precipitated particles into large units further apart [*bottom left on page 28*]. In such a region a dislocation would take on a wavy form bending around the centers that oppose its motion. But, as we have seen, a dislocation resists bending, so that now it is harder to move and the material is strengthened. Overly long heat treatment usually results in the state depicted in the third diagram, where the particles have conglomerated and separated so far that the long lengths of dislocation between them can bend comparatively easily and pass along, leaving dislocation loops surrounding each particle. Now the material is said to be overaged, and it has lost its hardness.



GRAIN BOUNDARY between contiguous particles of a polycrystalline metal is illustrated in the schematic diagram at top. The structure is equivalent to an array of edge dislocations as shown in the diagram at bottom.

Thus the dislocation theory has provided the first reasonable explanation for a number of time-honored but purely empirical metallurgical methods. But these successes are not the only evidence for the hypothesis. In the past few years more direct evidence for believing in dislocations has been uncovered.

In the first place the theory implies that a crystal without dislocations should be very strong. However, no one could figure out how to make such a crystal. Then Conyers Herring and J. K. Galt of Bell Laboratories found one in a faulty piece of telephone equipment. The tin in a certain capacitor was discovered to have grown tiny whiskers, about a fifty-thousandth of an inch in diameter, as the result of corrosion. These tin whiskers



AGE HARDENING takes place when the fine, closely spaced particles of a second structure (*diagram at top left*) are consolidated into larger particles as the result of aging (*bottom left*). If the process goes too far (*right*) the material is overaged.

ACID ETCH PITS, which occur at dislocations, outline a boundary between grains of a polycrystalline metal. The spacing between pits depends on the angle between the grains.

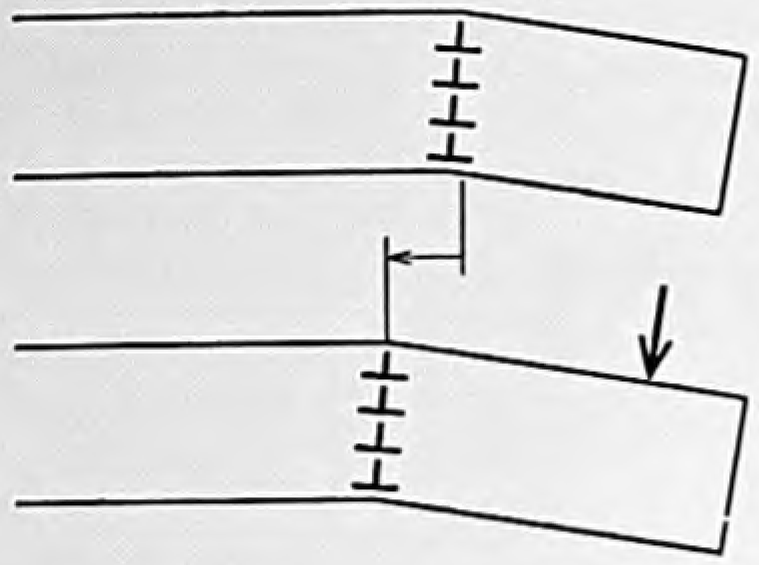
ruined the capacitor, but as Herring and Galt realized, they were worth far more than their weight in gold as research material. Within their tiny diameter there was no room for dislocations. When their strength was measured, they turned out to be nearly as strong as perfect crystals are supposed to be.

Another striking corroboration for dislocation theory concerns the boundary between adjacent crystal grains in a metal. Almost any piece of metal large enough to handle consists not of a single crystal, but of a large number of separate crystalline grains. The boundaries between these units are regions of misfit [*see boundary in bubble photograph on page* 25]. When the angle between grains is small, say less than 10 degrees, the boundary may be regarded as an array of edge dislocations, as shown in the diagram on page 27. Now it has been established that when a metal is treated with an acid, the rate of etching is higher in regions of higher energy. Therefore dislocation should be attacked more strongly than surrounding parts of the lattice. When a polished polycrystalline metal sample was treated with acid, the grain boundaries revealed a row of discrete etch pits or pips, each representing a single dislocation [*see photograph at left*].

The most convincing experiment thus far was performed in 1952 by E. R. Parker and J. Washburn of the University of California. Calculations indicated that when a metal is subjected to a shearing force, the rows of dislocations at grain boundaries should move in a direction perpendicular to the force. An ordinary polycrystalline sample is too complicated to analyze. Parker and Washburn grew two zinc crystals in a carefully controlled manner so as to produce a single low-angle grain boundary. This boundary behaved exactly as the theory predicted when a force was applied to the crystal pair.

Under some conditions the individual grains of a polycrystalline metal have been found to develop a fine network of subgrains. This state, called polygonization, usually results when a cold-worked metal is partially annealed (held at a moderately high temperature for a short time). Its effect is to strengthen the metal.

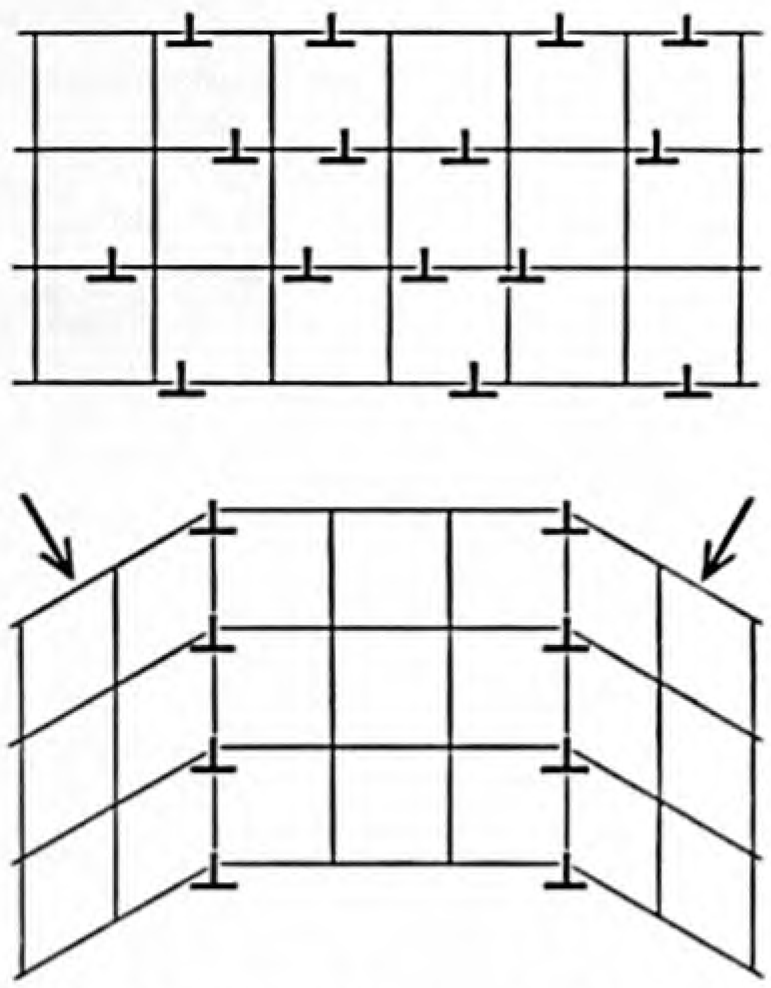
The mechanism behind this phenomenon can be seen in the diagram above. The first picture shows a metal lattice which contains a random distribution of dislocations. When such a metal is heated and deformed into an arc, the thermal agitation of the atoms makes the dislocations more mobile and they move to



BOUNDARY between a pair of large single crystals should move horizontally if a vertical shearing force (*arrow*) is applied as shown. Actual crystal boundaries have now been proved to behave as the theory predicts.

the more stable positions shown in the second picture. These positions are preferred because they involve the least distortion in the crystal. The process resembles what takes place when a thin, flexible steel ruler is bent. As the curvature increases, the ruler will suddenly spring into two straight sections forming a sharp angle—a more stable configuration under the imposed stresses. Since the polygonized structure represents a more stable position for the dislocations, they will now be harder to move. Hence the metal is stronger.

Metallurgy, one of the oldest arts, is only now coming into its own as a science. The dislocation idea is proving one of the most powerful tools of this new discipline. As yet its usefulness is chiefly theoretical, although the theory is already being applied in a few instances. For example, polygonization is a new method for strengthening certain metals. In general, however, it is beyond our powers to predict the actual behavior of the enormous number of interacting dislocations in a real metal. But as the theory develops we approach ever closer to the goal of a true understanding of metals and alloys.



POLYGONIZATION means the formation of tiny subgrains within each grain of a polycrystalline metal. It occurs when randomly spaced dislocations line up to give an arrangement of minimum strain in the lattice.

The Author

FRANK B. CUFF, JR., and L. McD. SCHETKY are both metallurgists at the Massachusetts Institute of Technology. Dislocations of a geographical sort played an important part in the youth of both men. Cuff, whose father was with the mining division of the Aluminum Company of America, received his early education at various stops in the U. S. and South America. Schetky, the son of a Navy officer, was born in the Philippines and went to school both in the U. S. and Europe. After graduating from the Rensselaer Polytechnic Institute, where he majored in metallurgy, Cuff received his doctorate from M.I.T. He is now an assistant professor at M.I.T., and is engaged in the study of creep and deformation in alloys at high temperatures. Schetky, also an R.P.I. graduate, took his Ph.D. there in 1953; a director of research at the Alloyd Research Corporation, he is presently working on the development of materials for use in ultra-precision instruments.

Bibliography

DISLOCATIONS IN CRYSTALS. W. T. Read, Jr. McGraw-Hill Book Company, Inc., 1953.

DISLOCATIONS IN METALS. Edited by M. Cohen. American Institute of Mining and Metallurgical Engineers, 1954.

DISLOCATIONS AND PLASTIC FLOW IN CRYSTALS. A. H. Cottrell. Oxford University Press, 1953.

IMPERFECTIONS IN NEARLY PERFECT CRYSTALS. Edited by W. Shockley and others. John Wiley & Sons, Inc., 1952.

SCIENTIFIC AMERICAN March 1952, Vol. 186, No. 3, pp. 47-54 OFFPRINT 205

THE QUANTUM THEORY

by Karl K. Darrow

To explain the radiation that comes from an enclosed cavity through a hole in its wall, Max Planck made an assumption whose consequences extended far beyond this problem, revolutionizing almost all of physics.

THIS is the history of a physical theory which began in 1900 by taking over a small province of physics and now has extended its empire over almost the whole of the sciences of physics and chemistry. More precisely, it is half of the history, for it carries the story only to 1923, which happens to be about as far along the road as the nonphysicist can travel easily. The story is that of the quantum theory, and this account will relate the main events in its early history.

Usually a doctrine or discovery in science does not spring unheralded and full-blown from a single man's brain; behind the reported founder of the doctrine there are likely to be others who partly anticipated him, and the origin of the idea may be lost in the mists. This cannot be said of quantum theory. The recognized founder was the actual founder, and there was no one behind him. His name was Max Planck; he was born in Germany in 1858, and he died there in his 89th year in 1947. Tacitus has said of some Roman character that he was *felix opportunitate mortis*—he was lucky to die when he did. Not so for Planck: he suffered grievously in the two world wars, losing a son in each and his house and his library in the second; he would have had a happier life if he had died in 1914. He would also have been equally famous, for his grand idea had come at the turn of the century.

Planck achieved his doctorate by a thesis on an experiment in the diffusion of hydrogen through palladium. This was the only experiment he ever performed: one may suspect that it was imposed on him rather than chosen by him. Even before he won his doctorate, he had devoted himself to the theoretical study of the foundations of thermodynamics. He saw rather more deeply into the second law of thermodynamics than anyone before him, and his textbook on thermodynamics is still a classic.

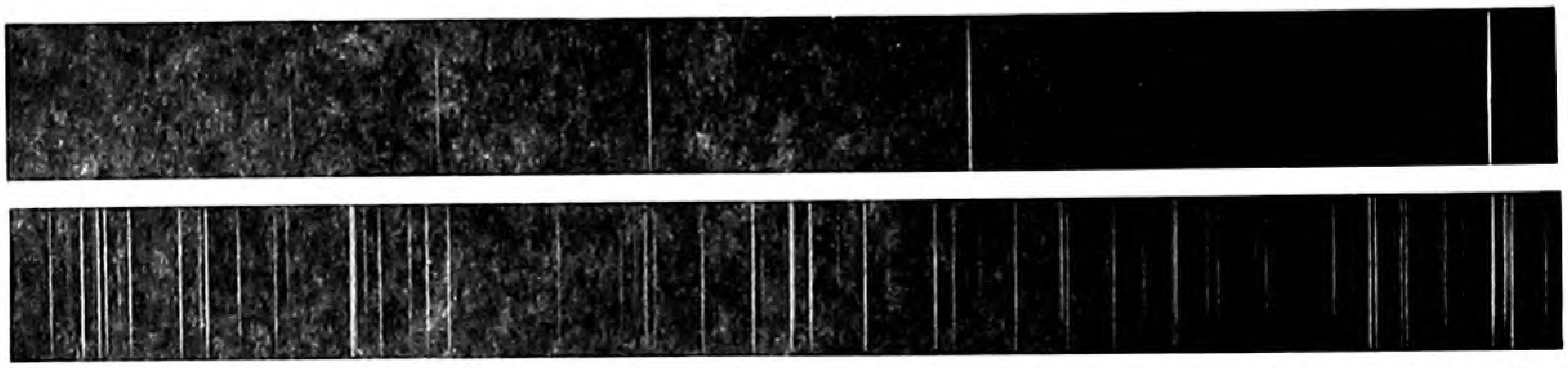
It was by way of thermodynamics that Planck came to the quantum theory. Thermodynamics is a hard subject, and Planck arrived at the quantum theory by the hard way. When he entered upon thermodynamics, it was a science of heat in matter. Planck extended it to light.

The Cavity Problem

Imagine a cavity inside a solid body. There is a small perforation in the wall of the cavity, out of which light can come. It must be large enough so that we can get enough light out of it to analyze with a spectroscope; it must also be small enough so that the light which comes out will be a fair sample of what the light inside would be if the perforation were not there. We have therefore to assume that one and the same perforation can be neither too large nor too small. This sort of double-barreled assumption is frequently made in physics, and in this case at least it has not yet led us into any trouble.

First let us suppose that there is some gas inside the hollow. For definiteness let the gas be helium, which is a nice monatomic gas and nearly a perfect gas. The gas has energy, which is the kinetic energy of its atoms. This kinetic energy we know to be proportional to the absolute temperature. We also know that the energy is proportional to the number of atoms. If I pump half of the gas out of the cavity while the temperature remains the same, the energy of the gas in the hollow is reduced to one-half; if I pump extra helium into the cavity so as to double the quantity, I double the energy.

Let us now imagine that all of the gas is pumped out. The cavity is still not empty, for it still contains light, in fact, just the same amount of light as while the gas was there. Since the cavity contains light, it contains the energy of that light. This energy is proportional, however, not simply to the temperature but to the *fourth power* of the temperature. There is also a much more important difference between light and helium: namely, whereas the energy of the helium in the cavity is proportional both to the absolute temperature and to the amount of helium, the energy of the light depends only on the temperature. There is no point in saying that the



SPECTRA OF HYDROGEN show bright lines which correspond to energy transitions in the hydrogen atom and molecule. The brighter lines in the top spectrum are emitted by atomic hydrogen; the more complex lines in the bottom spectrum, by the hydrogen molecule. The spectra were made at The Johns Hopkins University.

energy of the light is proportional to the amount of light, for the amount cannot be varied. There is no pump by which I can pump out half of the light and leave the rest. I could indeed increase the amount of light in the hollow by shining a searchlight through the perforation, but this would make no lasting difference. The beam of the searchlight would be absorbed in the walls of the cavity and heat them up; as soon as I turned off the searchlight, the temperature would start to revert to its original value, and when it got back to its initial value there would be the same amount of light in the cavity as before.

The light emerging through the perforation has a perfectly distinctive spectrum, and this we have already assumed to be identical with the spectrum of the light inside the cavity. On a graph in which energy is plotted against frequency, the spectrum has a single broad hump, meaning that the energy is greatest at the middle wavelengths in the spectrum and least at the highest and lowest frequencies. This spectrum has several names. Planck called it the "normal spectrum." Germans generally call it the spectrum of "cavity radiation," which is the best name. English-speaking people usually call it the "black-body spectrum," because a perfectly black body would emit light of this composition. (It seems paradoxical that a black body should emit light, but it is no paradox at all when we remember the distinction between emission and reflection. The practical definition of blackness is that it is the quality of not reflecting light. A black body does not reflect, but it may emit; in fact, it is the most powerful possible emitter of light. We ordinarily look at bodies—black, white or colored—when they are reflecting light but are not hot enough to emit an appreciable amount of visible light. At temperatures of incandescence a black body would look brightest of all.)

Little Resonators

Planck set out to explain the black-body spectrum. He postulated that all solid bodies, and therefore the walls of the cavity, contained little oscillators—he called them resonators—of every imaginable frequency. By resonators Planck meant small electrified particles which wiggled to and fro. (A few years later he identified them as vibrating electrons, but in 1900 the electron had only just been discovered and was far from well established in physics.) It followed that in any range of frequencies, however narrow, there must be resonators having frequencies in that range. Take, for instance, the bright yellow spectrum-line of sodium which is designated by the letter D. We are to imagine that there are resonators having frequencies within this narrow range. All of them are emitting yellow light into the cavity, and yellow light is coming right back out of the cavity and bathing them and stimulating them to continued oscillation. There is an equilibrium between the energy of the resonators and the energy of the yellow light. So it is with the energy of blue light and the resonators which have the frequency of blue light, and so it is throughout the entire spectrum. If by any valid line of reasoning one could arrive at the energy which the resonators of any frequency have when they are in equilibrium with the light, one would be able to arrive by one simple further step at the energy which the light of any frequency has when it is in equilibrium with the resonators. This would be the answer to Planck's question, for it would give the black-body spectrum.

There was an answer to this question before Planck, but the answer was a disaster. This "classical" answer was that all of the resonators have the same energy, no matter what their frequency. From this it followed that the quantity of light in the cavity would be infinite, and since the quantity of light in the cavity actually is finite, the theory was sunk.

The flaw in the classical theory lay in one of its assumptions—indeed, a very plausible one—to wit: that any resonator may have any amount of energy; in other words, that it may go continuously up or down the energy-scale through all levels, like a man walking up or down a ramp. Planck replaced the ramp by a ladder, with its rungs, as in any practical ladder, at equal intervals. His assumption was that the resonator must always be on one or another of the rungs, and if it proceeds up or down the ladder, it must proceed by jumping from rung to rung. That is to say, it may absorb or emit energy only in certain discrete units, represented by the equal spacing from rung to rung of the ladder. Planck then affirmed that the spacing between the rungs of the ladder was not the same for all resonators but must be proportional to the frequency of the resonator in each case. He designated the spacing, or energy unit, as $h\nu$, and said that h is a universal constant, ν being the frequency of the resonator.

Planck's substitution of the ladder for the ramp not only avoided the disaster of infinity but yielded the actual form of the black-body spectrum. To explain just why it did so, it would be necessary to define entropy; to quote what is called Boltzmann's relation between entropy and probability; to define probability in a very special and peculiar way; to derive the formulas for the probability and the entropy of a flock of resonators, among which parcels of energy of the amount $h\nu$ are distributed in the most probable way; to say that the derivative of the entropy with respect to the energy is the reciprocal of the absolute temperature; to show that when all this is done, one arrives at a formula which gives the energy of the resonators of any chosen frequency ν as a function of the temperature, and to conclude by showing that out of this formula comes the black-body spectrum which is ratified by experiment. This skeleton of the argument shows that the argument itself can find no place in an article designed for people who are not mathematical physicists. It will certainly be conceded that Planck got to the quantum theory the hard way.

The constant h is known as Planck's constant. Its value must be determined by experiment; it is not prescribed by theory. It has been measured in a variety of ways, and all of the measurements agree. Planck may be characterized as the man who put h into physics. People who drop the h are sometimes called cockneys; in this sense the physics of the period before Planck may be called cockney physics. If the physicists of that period could return for a glimpse of our physics, they might say that it is cockeyed physics—and yet ours is right, and they would be wrong.

I have said that the rungs of the ladder—the "levels," as they are more generally called—are spaced at intervals $h\nu$; I have not said where the ladder starts. In Planck's original theory the bottom rung was at zero: the energy-values permitted to a resonator were 0, $h\nu$, $2h\nu$, $3h\nu$, and so on. Later Planck revised his theory and located the bottom rung at the energy-value $(\frac{1}{2})h\nu$, so that the subsequent rungs followed along at $(\frac{3}{2})h\nu$, $(\frac{5}{2})h\nu$, and so on. So long as we concern ourselves with black-body radiation alone, it makes no difference where we start, but other phenomena confirm Planck's second choice.

Particles of Light

If light is emitted and absorbed only in units of size $h\nu$, the obvious next step is to infer that light travels around in units or parcels or particles of energy $h\nu$. But this is a step that Planck himself did not take. According to some of those who knew him and to the evidence of his own writings, Planck was a revolutionary of a very conservative kind. He took one radical step when it was necessary, but he was not daring enough to try to revive the corpuscular theory of light, and this is understandable in view of its state at the time.

In 1900 the corpuscular theory of light was very dead. Newton's idea that light consisted of particles had yielded in the 19th century to the wave theory, and almost anyone who might have attempted to restore the corpuscular theory would have run grave danger of being rated a crank. To resurrect it there was needed a man possessed of (a) prestige, (b) courage, (c) knowledge

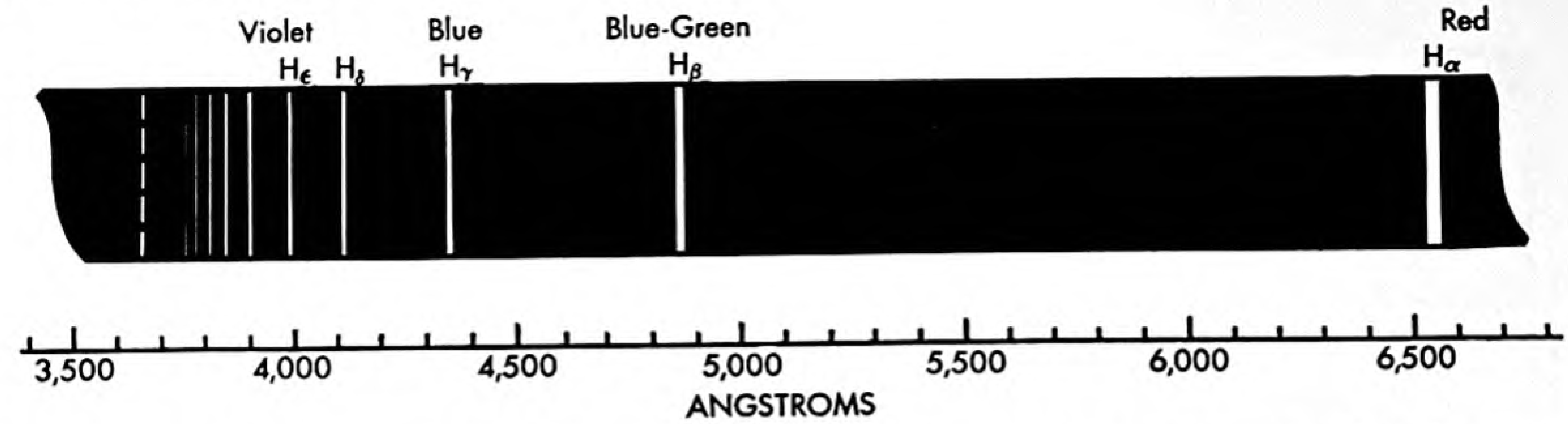
MAX PLANCK was born in 1858 and died in 1947. On the wall behind him in this only partly fanciful reconstruction of his study are the portraits of nine men of notable importance in his scientific life. **From left to right they are: Helmholtz, Einstein, Bohr, Clausius, Pringsheim, Lummer, Rubens, von Laue, Sommerfeld.**

of Planck's theory and (d) knowledge of certain experiments which had not yet been performed when Planck put forth his theory. There was such a man, and his name was Einstein.

The new experiments were those on the photoelectric effect. This effect consists in the emission of electrons from metals under the influence of light. Think of a beam of light playing on a piece of metal—sodium, for example. The beam of light, Einstein imagined, consists of corpuscles of energy $h\nu$. Some of them are reflected from the surface of the metal; these do not concern us. Others enter the metal, and one of these may be absorbed in such a way that it vanishes completely, yielding up its entire energy $h\nu$ to a single electron in the metal. The electron thereby gains enough speed and energy to spring out of the metal, and it becomes a "photoelectron." Once it is out, we can measure its kinetic energy.

This energy of course is something less than $h\nu$. The electron has had to spend a certain amount of its energy in order to leave the metal; we might describe this as the fee it pays at the surface of the metal for its exit-visa. We denote this fee by the letter W; thus the energy the electron has left when it gets out to where we can observe it is $h\nu$ minus W. If we could measure only photoelectrons released by one frequency of light we should be unable to separate W from $h\nu$. But we can experiment with beams of light of many different frequencies, and can find the kinetic energy of electrons emitted under the impact of each frequency. The theory says that if we plot the electron-energy E against the frequency ν, the points should lie along a straight line, described by the equation: $E=h\nu-W$. Experiments confirm that the points *do* lie along a straight line, and that the slope of this line is h, Planck's constant. As for W, it can be determined too; but we will leave that constant to the people who are most interested in it—the students of solid-state physics.

The discovery of this equation and the revival of the corpuscular theory of light are milestones in the history of physics. Einstein's theory inspired many famous experiments; one of the chief experimenters was Robert A. Millikan. It must be remarked that even Einstein was not bold enough to come out for the corpuscular theory of light unreservedly. He qualified his idea with a singular word, which I think proper to call a hedging word—"heuristic." According to the *New International Dictionary* heuristic means "serving to discover or reveal—applied to methods of demonstration which are persuasive rather than compelling, or which lead a person to find out for himself." I interpret that Einstein meant to convey that light acts in the photoelectric effect as though it



SPECTRAL LINES of atomic hydrogen are mapped against their color (*top*) and wavelength in Angstroms (*bottom*). This is the celebrated Balmer series, one of several such spectral series which reflect the energy transitions of the hydrogen atom. The study of these series led to Bohr's model of the atom (*see next page*).

were corpuscular, without committing himself further.

Today there is no longer any reason to hedge. The corpuscular theory of light has been re-established. In its reincarnation it is the first child of the quantum theory. The art by which it is combined with the wave theory is also a triumph of quantum theory, but this is a part of the later history. A remark about language is appropriate here. It must come as quite a surprise to a student of language to learn that "quantum theory" and "quantum mechanics" and "quantum number" are terms very common in physics, while the word "quantum" itself has practically vanished. Quantum, first introduced to denote the quantity $h\nu$, later came to be applied to the corpuscles of light. In this latter sense it has been superseded by a much better word—photon.

Vibrating Atoms

The next province of physics to be invaded by the quantum theory was the province of heat in matter. The cavity exhibits heat in the form of radiant energy. Heat in solids is identified with the vibrations of the atoms. Since Planck had scored quite a resounding success with the assumption that oscillators interchange energy with radiation in discrete units, nothing could be more natural than to try out the same idea on the oscillating atoms in solids. Like other ideas which seem very natural after somebody else has thought of them, this idea took several years to emerge. Again we must salute Einstein as the theorist, and on this occasion Walther Nernst of Germany was the leading experimenter.

The classical theory (note that in physics the term "classical" is often applied to the last-but-one theory in the field) said that the heat content of a solid is proportional to its absolute temperature. Another way of stating the theory is that the specific heat of a solid, which is the amount of heat needed to impart a unit increase in temperature to a unit mass of the substance, should be independent of the temperature at which the solid happens to be when the heat is applied. Actually the amount of heat needed to raise the temperature of a solid by a unit (say one degree) falls off as its temperature declines. As experimentalists made progress toward lower and lower temperatures, the departures from the classical theory became more and more striking. The quantum theory arrived in time to take care of them. It was found that the ladder of levels is the same for the vibrating atom as for Planck's electrical resonator. Since there is no rung below $(\frac{1}{2})h\nu$, the quantum theory tells us that a vibrating atom can never have a lesser energy than this—not even at absolute zero temperature, where according to classical theory the atoms should be standing absolutely still. This singular result of quantum theory has been attested by experiment.

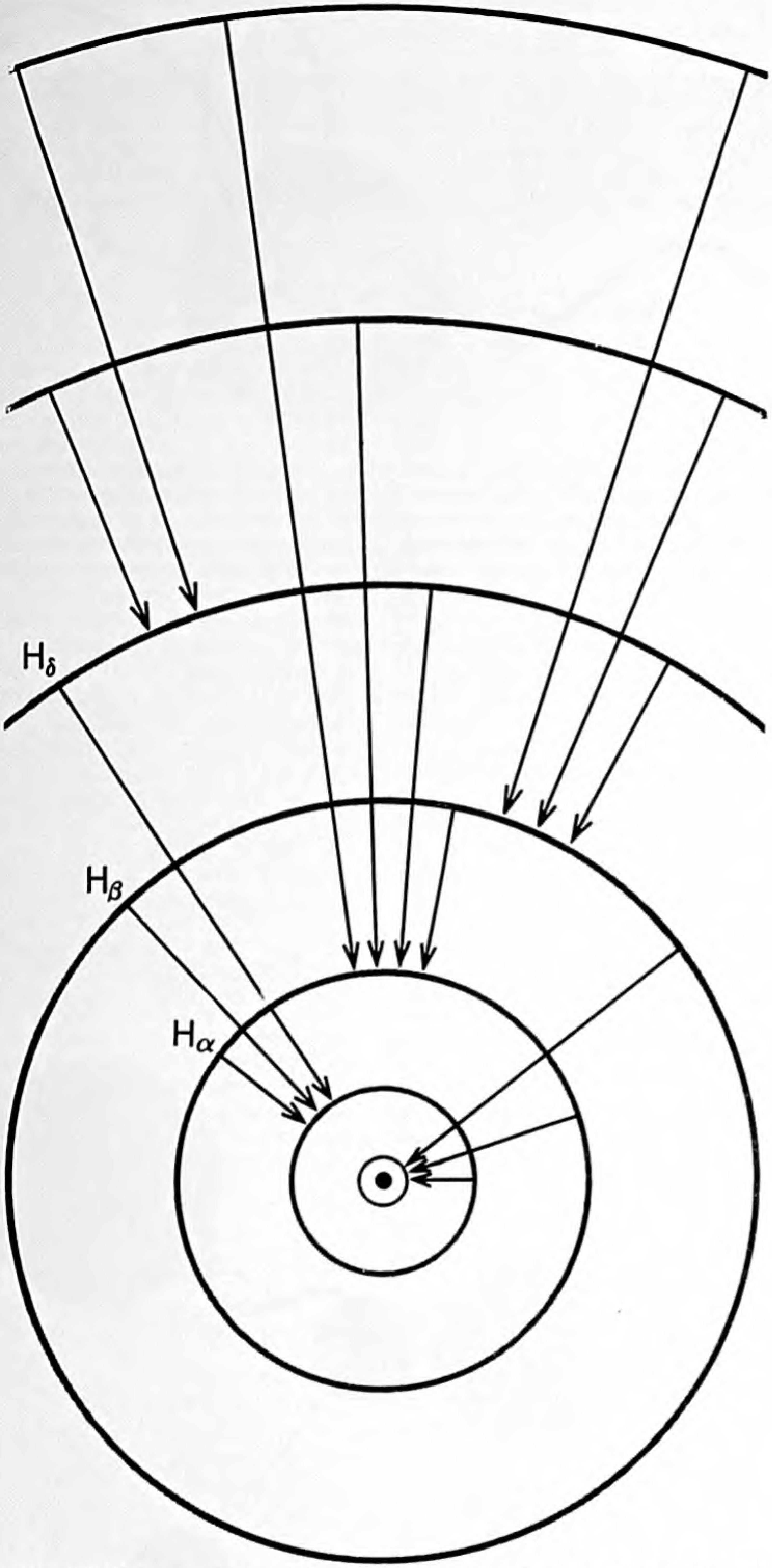
Another type of substance which defers to quantum theory is the gas whose molecules contain two atoms, such as hydrogen, H_2. The energy picture of a monatomic gas such as helium is simple: the kinetic energy of its atoms is purely the energy of traveling motion (*i.e.*, without the complication of rotation of the atoms themselves), and the quantum theory has nothing new to say about it—not at least until the temperature falls extremely low. But the molecules of a diatomic gas not only move about as traveling bodies but also rotate around the axis connecting the two atoms—think, if you will, of a dumbbell spinning end over end. The heat content of such a gas is partly the kinetic energy of traveling motion and partly the kinetic energy of rotatory motion. It was no slight step forward to imagine that rotatory energy may be limited to definite energy-values just as vibratory energy is. Starting from this advance, the quantum theory, which had already extended its domain over the fields of black-body radiation, the photoelectric effect and specific heat, was now to conquer the atom.

Before we begin that part of the story, I should make clear what is meant by the statement that rotatory energy is limited to certain definite discrete values, or, as the physicist says, is "quantized." Rotatory motion involves the concept of angular momentum, one of the most important in physics. For a particle revolving in a circular orbit, the angular momentum is the mass times the angular velocity; for a rigid body rotating around an axis passing through itself, the angular momentum must be found by calculation; often in quantum physics it appears as a quantity which is easy to measure but impossible to separate into mass and angular velocity.

The law of quantization for angular momentum is the simplest of all laws. In this ladder the unit of spacing between the evenly spaced rungs is $h/2\pi$, and the scale of permitted values of angular momentum usually runs: 0, $h/2\pi$, $2h/2\pi$, and so on. It may be surprising to find Planck's constant h reappearing in this new role. First it was a factor by which the frequency of a vibrator must be multiplied in order to get the spacing between the permitted energy-values; now, multiplied by $1/2\pi$, it is the interval between the permitted values of angular momentum of a rotating body. It can play both these roles because it has what physicists call the correct dimensions for both. Planck's constant has the dimensions of energy divided by frequency, and these are also the dimensions of angular momentum.

Bohr

The quantum theory's conquest of the atom began in the year 1913 (remembered with nostalgia by those who are old enough as the last year of the good old times), and the leader in this conquest was a young Dane by the name of Niels Bohr, who came to the University of Manchester to study with the celebrated Ernest Rutherford. Rutherford had just established the conception that the atom is a sort of miniature of the solar system, in which the planets are represented by negative electrons and the central sun by a positively-



BOHR MODEL postulated a hydrogen atom in which an electron traveled around a nucleus in any one of several "permitted" orbits. When the electron fell from an orbit of higher energy to one of lower energy, it emitted light of a characteristic wavelength. In this diagram the electron orbits are indicated as circles or sections of them. Each group of arrows represents a group of energy transitions which give rise to a spectral series. The three labeled arrows at the left refer to the Balmer series shown on the preceding page. The labels correspond to three spectral lines in that drawing.

charged and massive nucleus, around which the electrons revolve in orbits as the planets revolve around the sun. In the simplest case, the hydrogen atom, there is only one electron. Rutherford did not prescribe the particular orbit in which this electron should revolve. To prescribe it required a revolutionary new idea, and Bohr was the man who provided this idea.

The orbit of the electron around the nucleus, like that of a planet around the sun, should be an ellipse, for in both cases the force is an inverse-square force. A circle is a particular case of an ellipse; we will consider only circular orbits. A hydrogen atom, with its electron revolving in a circular orbit about its nucleus, can be regarded as a wheel. It is a peculiar kind of wheel, since it has no spokes and the rim is vacant except for the small region occupied by the electron, but it possesses the major property of a wheel: angular momentum. Bohr made the assumption that the angular momentum of a hydrogen atom is quantized, and that the electron is permitted to revolve in certain orbits which correspond to the integer multiples of $h/2\pi$. In its "normal state" the hydrogen atom has its electron revolving in the circle for which the angular momentum is $h/2\pi$. Its other permitted orbits are those for which the values of the angular momentum are $2h/2\pi$, $3h/2\pi$, and so on. Each of these other orbits, Bohr postulated, represents an "excited state" of the hydrogen atom.

The energy of the hydrogen atom in its normal state is $-R$; R being a constant of which Bohr computed the theoretical value. The general formula which expresses the various energy-values permitted to the hydrogen atom is $E_n = -R/n^2$. In its first excited state, represented by $n=2$, its energy is $-R/4$; in the second excited state it is $-R/9$, and so on. One may be puzzled by the fact that all of these energy-values are negative. This is because we are reckoning energy from a zero which corresponds to the state in which the hydrogen atom is completely torn apart, with the nucleus and the electron infinitely far from each other. It would seem more rational to reckon energy from a zero which corresponds to the normal state of the hydrogen atom, but this actually makes the formula more complicated.

Bohr went on to propose a second revolutionary idea, without which the first would have been of little use. Imagine a hydrogen atom in, let us say, the second excited state—the one for which the energy is $-R/9$. Suppose that the electron transfers itself into the orbit corresponding to the first excited state—the one for which the energy is $-R/4$. The atom now loses the difference in energy $(R/4 - R/9)$. What happens to this energy? According to Bohr's second idea, it leaves the atom, in the form of a

single photon. Accordingly, there should be, in the spectrum of hydrogen, a line of which the frequency is equal to $(R/4 - R/9)/h$. Moreover, there should be lots of other lines, and all of their frequencies should be calculable by inserting various integer values for m and for n in the general formula:

$$\nu-(R/m^2 - R/n^2)/h$$

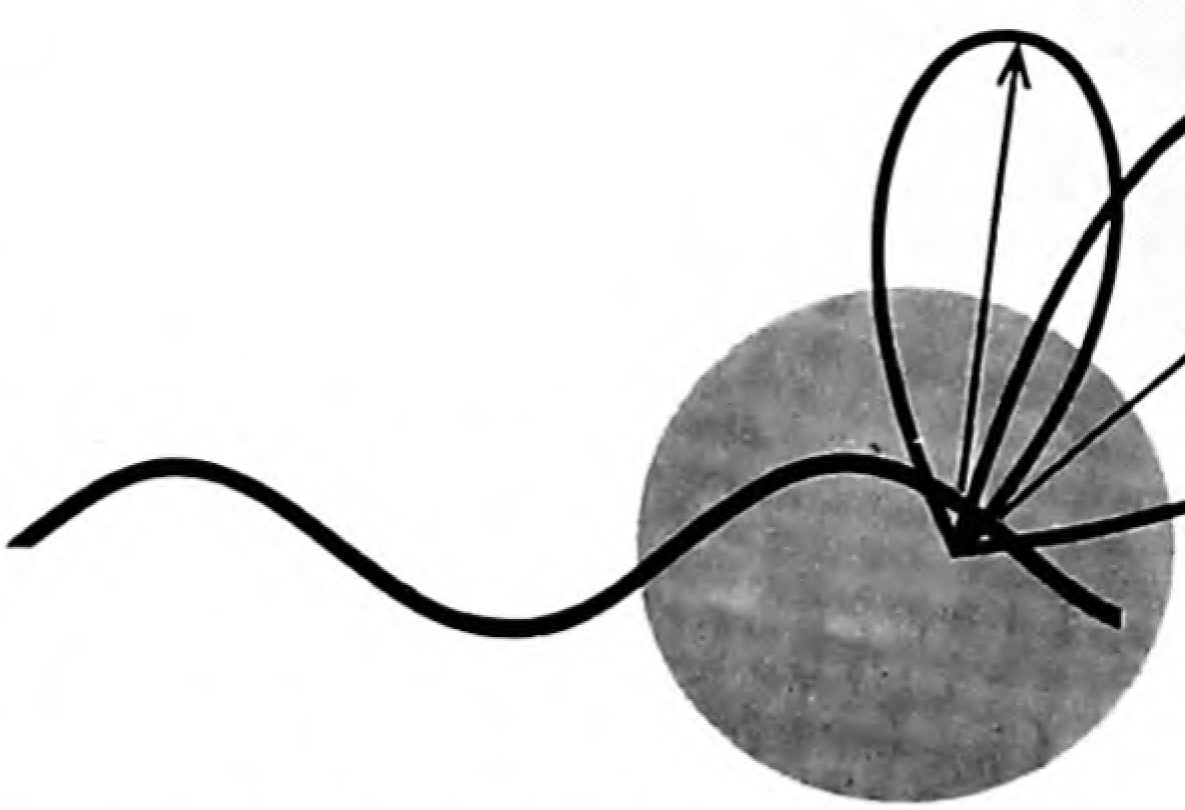
Well, this is a correct description of the actual spectrum of hydrogen, and the theoretical value assigned by Bohr to the constant R agrees with the value derived from experiments. There is nothing in the world which impresses a physicist more than a numerical agreement between experiment and theory; and I do not think that there can ever have been a numerical agreement more impressive than this one, as I can testify who remember its advent.

This is a good place to interpolate that one of the most curious features of quantum theory is that one and the same formula can be derived from markedly different postulates. The picture of the hydrogen atom today differs somewhat from the one Bohr proposed, and yet the new picture leads to the same formula for the spectrum of hydrogen and to the same value of R. The conjurers of quantum theory get the same rabbit out of more than one hat.

Spins

Now let us consider some other wheels. The first of these is the electron. The electron, besides revolving around the nucleus, possesses an angular momentum of its own, and we liken it therefore to a wheel. It may be visualized as a rigid body spinning upon its axis, but this is a rather dangerous analogy, for it leads one to inquire what the electron's angular velocity is, and no one has ever been able to answer this question—indeed, it is very likely unanswerable. The electron's angular momentum is quantized, and in the simplest conceivable fashion: this is a ladder with only one rung. The angular momentum of the electron is fixed forever at the single value $(\frac{1}{2})h/2\pi$.

The electron is considered to be an elementary particle; I call it a "structural elementary particle," meaning a particle which is used in our models of atoms and of the nuclei of atoms. There are two other structural elementary particles: the proton and the neutron. The proton is the nucleus of the commonest and lightest kind of hydrogen atom. The neutron cannot serve as the nucleus of an atom when it is by itself, but it combines with protons and with other neutrons to form composite nuclei. These three structural elementary particles have one quality in common, and only one. Each has the same one-rung ladder of angular momentum; each has the unalterable angular momentum $(\frac{1}{2})h/2\pi$.

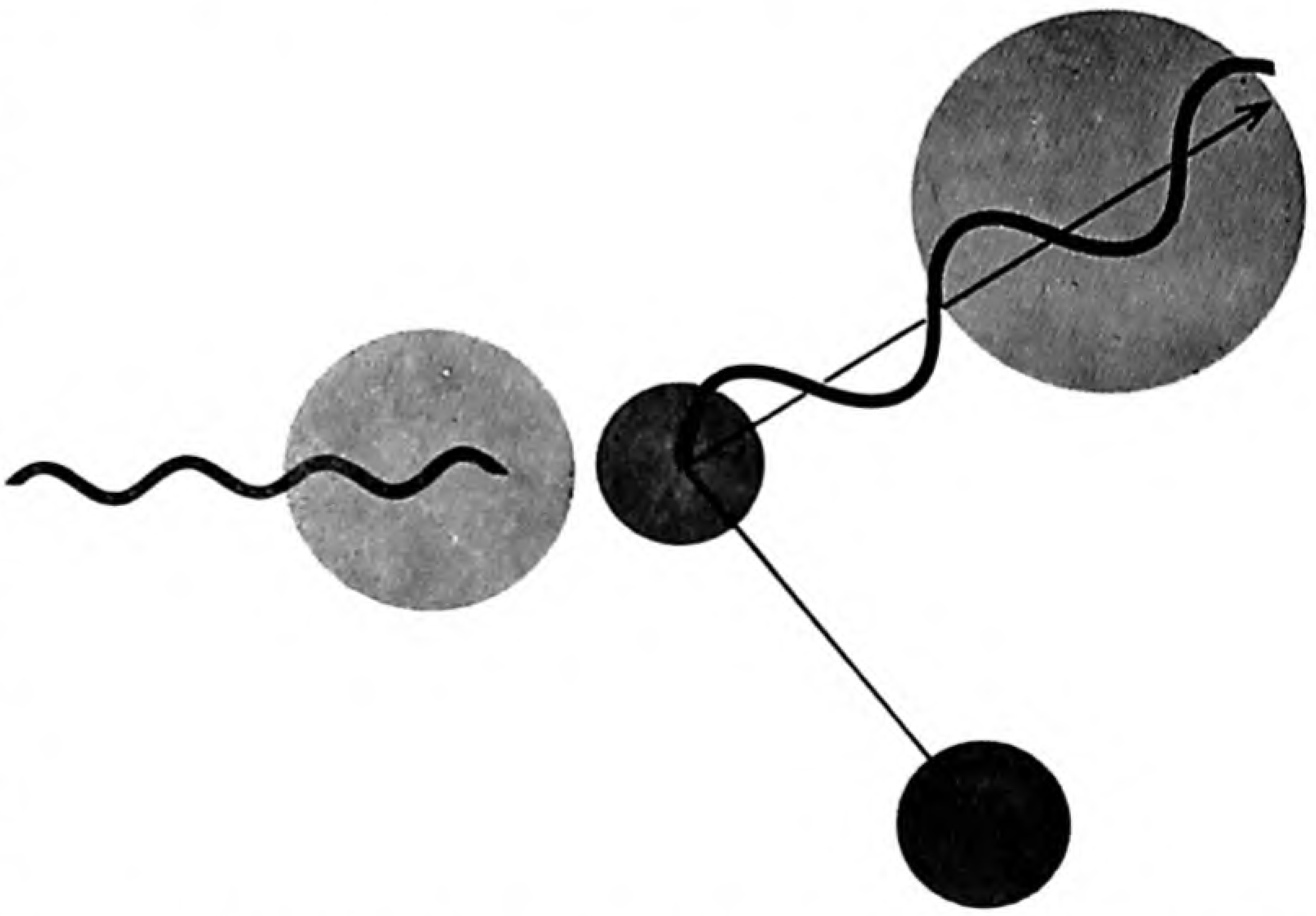


PHOTOELECTRIC EFFECT occurs when a photon (*wavy line*) encounters an atom (*gray circle*) and knocks out one of its electrons (*arrows*). A low-energy photon causes the electron to leave the atom at a steep angle; a high-energy photon, at a shallower angle. The curves around the arrows show the angular distribution of electrons resulting from many such photons.

The most important feature of angular momenta, on the atomic and the subatomic scale, is the so-called law of composition, meaning the composite angular momentum of a system consisting of two or more particles. I will illustrate this law by three examples.

The first is the common light hydrogen atom, containing one proton and one electron. The spins of these two particles are always either parallel or antiparallel to each other. Thus there are two, and just two, kinds of light hydrogen atoms. The difference in their properties is very slight indeed, and yet it can be detected by a spectroscope operating in the microwave region.

My second example is afforded by the hydrogen molecule, consisting of two common hydrogen atoms. This has two protons and two electrons. The two electrons always have their spins pointed antiparallel to each other. The two protons may be pointed in the same or opposite directions to each other; there are no intermediate cases. Thus there are two, and just two, kinds of molecular hydrogen. Their physical properties differ appreciably, and they can be separated from each other. They even have



COMPTON EFFECT occurs when a high-energy photon (*wavy line and gray circle at left*) encounters an electron (*smaller circle in center*). The electron is knocked away at an angle (*lower right*). The wavelength of the photon is decreased by the amount of energy imparted to the electron (*upper right*).

different names: ortho-hydrogen and para-hydrogen.

My last example is provided by the simplest of all composite nuclei—the deuteron. This is the nucleus of the isotope which used to be called heavy hydrogen, though this name is waning now that a still heavier isotope is known. The deuteron consists of one proton and one neutron. In every deuteron the spins of the proton and the neutron are parallel. This is a sharper restriction than prevails in my other two examples. The theorists believe that the spins of the proton and the neutron have no objection *per se* to setting themselves antiparallel, but that the forces between these two particles are such that in the antiparallel orientation the proton and the neutron just cannot stick together.

Notice now what would be the consequence if this law of composition of angular momenta did not exist. There would be not just two kinds of hydrogen atom but an infinite variety of kinds, though it is true that they would not differ appreciably in any significant quality. There would be not just two kinds of molecular hydrogen but an infinity of different kinds, with properties lying all the way between those of parahydrogen and those of ortho-hydrogen; these would spoil the distinctiveness of molecular hydrogen. There would also be a wide variety of kinds of deuteron.

Follow this idea a step further. Suppose that the electron of the hydrogen atom could revolve around the nucleus in any orbit whatsoever, and not just in a limited number of orbits of which one corresponds to the normal state and the others to transient excited states. Were this true, there would be an infinity of different kinds of hydrogen atoms. This in turn would mean that hydrogen would not be the distinctive and individualistic element that it is. The same may be said about every other element. But for the laws of quantization, carbon would not be carbon as we know it, oxygen would not be oxygen, iron would not be iron and gold would not be gold. What the quantum theory explains is the distinctiveness and the individuality of the 90-odd elements of which the world is made. What a long way for a theory to have come, that started out as a theory of the recondite subject of black-body radiation!

Photon Collisions

I will mention one more attribute of light into which Planck's constant has made its way. It concerns a concept which would have been inconceivable in 1900 and was not as a matter of fact conceived until 1923: namely, that a photon, or corpuscle of light, is so much a particle that it can have an elastic collision with an electron. This phenomenon was observed and interpreted by Arthur Compton, and it is therefore known as the Compton effect.

An elastic impact between two bodies is one in which both kinetic energy and momentum are conserved. After the impact each body has a different energy and a different momentum from what it had before, but the sum of the energies of the two bodies, as well as the sum of their momenta, is still the same. What Compton did was not only to propose that such an impact could occur between a photon and an electron but to state a formula for the transfer of energy and momentum.

If a corpuscle of light has linear momentum, as is implied by the fact that light exerts pressure, the momentum of a photon of frequency ν should be $h\nu/c$, the symbol c standing for the velocity of light. Consider an elastic impact between a photon and a stationary electron. The initial energy and the initial momentum of the photon are $h\nu$ and $h\nu/c$, respectively; since the electron is stationary, its initial energy and momentum are zero. Let us suppose that the electron recoils from the impact in a direction at a certain angle with the direction in which the photon was originally traveling. Now the equations of the impact are easy to solve. One can calculate the angle of recoil of the photon, its new energy, which turns out to be less than its original energy $h\nu$ and its new momentum, which turns out to be less than its original momentum $h\nu/c$.

We can now write an equation which gives the new frequency of the photon: its value is E'/h—the new energy divided by Planck's constant. In short, what this equation and the foregoing theory say is that if the electron recoils in a certain direction, the photon goes off in a certain calculable direction and has a lesser frequency (or longer wavelength) than it had before, this new frequency also being calculable.

To test this theory, it is only necessary to set up an X-ray spectroscope to catch the photons going off in the theoretically calculated direction. When this test was actually made, by directing X-rays (photons) against a target of matter rich in electrons, Compton's theory was completely confirmed.

Now we have reached the twin climaxes of the early quantum theory: the law of composition of angular momenta and the Compton effect. We have also reached the year 1923. At this point in the story quantum theory suffers a mighty change. Its expression in terms of mathematics becomes much harder; its expression in terms of words and of analogies with concepts of the past becomes so very much harder as to verge on the impossible. But if one is interested only in the simpler applications of the theory, it is not necessary to attempt to pass through the formidable portal.

The Author

KARL K. DARROW, now retired, was for many years a physicist at Bell Telephone Laboratories and secretary of the American Physical Society. He was born in Chicago in 1891 and educated at the University of Chicago, taking his doctorate in 1917 and then going on for postgraduate studies at the Universities of Paris and Berlin. From 1917 to 1956 he was theoretical physicist at Bell Telephone Laboratories, holding at various times visiting professorships at Stanford University, Chicago University, Columbia University and Smith College.

Bibliography

THE NEW CONCEPTIONS OF MATTER. C. G. Darwin. The Macmillan Company, 1931.

SCIENTIFIC AMERICAN June 1949, Vol. 180, No. 6, pp. 30-39 OFFPRINT 206

LOW TEMPERATURE PHYSICS

by Harry M. Davis

The ceaseless motion of atoms and molecules would be stilled at absolute zero. Workers who have reduced matter close to this state observe phenomena that enrich our understanding of the physical world.

IT was quitting time at the National Bureau of Standards. Most of the technicians were joining their car pools to enter the stream of traffic on Connecticut Avenue, but in the low-temperature laboratory a fresh batch of the coldest fluid was ready to be siphoned from Russell Scott's home-made helium liquefier. It was J. R. Pellam's turn to draw the precious refrigerated liquid, and dinner could wait. As he set the vacuum-insulated flask of liquid helium carefully inside another vacuum-insulated flask of liquid oxygen, connected the vapor-pressure thermometer and adjusted a clamp on the hose to the vacuum line, Pellam made a remark that explained much of the fascination that low-temperature physics, or cryogenics, has for its workers. "We are beginning to learn," he said, "what lies beyond the thermal chaos."

Thermal chaos is more than a figure of speech. It is the random movement of molecules in every substance we touch, in the air around us, in our very protoplasm. It is thermal chaos, in fact, that makes life possible. Without it molecules could not collide and interact in the ceaseless round of metabolism.

One of the greatest discoveries of the 19th century was that heat was not a "subtle fluid," as earlier philosophers had thought, but a condition of matter. In England, where thermal chaos was first harnessed to drive the steam engine, scientists began to pronounce in their lectures that "heat is a mode of motion." It was a triumph of this period that the laws of the pressure, the volume and the temperature of gases could be drawn from the statistics of vast numbers of particles moving and colliding in a fortuitous fashion. The totality of their impacts pushed the moving piston; the totality of their kinetic energies was their content of heat.

Normally we experience the smooth, regular result of the law of averages working over enormous populations of moving molecules. We have the steady reading of a thermometer, the smooth spinning of a steam turbine, the definite speed of sound in air of a given temperature. Under certain conditions, however, the thermal chaos can be seen and heard. It can be seen in the microscope, under which, as the British botanist Robert Brown observed more than a century ago, tiny particles are never entirely still. Brown's name is perpetuated in this Brownian movement.

It is easier to listen to thermal chaos. The only necessary apparatus is a radio set. The noise heard when the set is not tuned to a station and the volume is turned up is nothing but the amplified effect of electrons boiling at random in the thermionic tubes. In fact, the thermal agitation that the power engineer defines as heat has a less flattering definition in the vocabulary of the communication engineer. It is "pure noise," the frustrating factor that limits the sensitivity of a receiver, since more amplification will build up the set's own electronic noise as much as the incoming signal. One of the practical applications that can now be envisaged from low-temperature research is the use of a refrigerated crystal, free of internal noise, to pick up radio signals far fainter than the threshold of a heated radio tube.

Accepting the fact that matter at familiar temperatures is in a state of thermal chaos, there are two ways of bringing order out of it. The usual way is to deal with vast molecular populations, where the laws of probability work out so that individual fluctuations are not noticeable. This, perforce, will continue to be the usual way of doing business with nature. The other way, which is pursued in cryogenic laboratories, is to remove as much as possible of the energy of motion, silence the "pure noise" of random movement, and see what happens when matter approaches utter stillness.

We shall see that below the temperatures of thermal chaos matter behaves in strange and excitingly different ways, exhibiting novel responses to the stimuli of electricity, magnetism and heat. There are superconductors of electricity, screens against magnetism, new forms of wave motion, and, in the case of helium, a "fourth state of matter" which cannot be strictly defined as either a liquid, a solid or a gas. These odd phenomena have made low temperatures one of the most fascinating frontiers of current physical research.

Most U.S. work at this frontier is sponsored by the Office of Naval Research, the scientific administrators of which have not demanded immediate practical results, military or otherwise. It is enough for them that there is knowledge to be gained that will lead to a better understanding of metals, crystals, liquids and gases, of electrical resistance

and induction, of electrical conductors, semiconductors and superconductors. Asked about the usefulness of it all, they will refer to the classical cliché, "What good is a newborn baby?"

Low Temperatures in Industry

We might first examine some of the baby's older brothers, earlier progeny of low-temperature research that have grown to healthy maturity. On the long road from zero degrees Centigrade (the freezing point of water) to zero degrees Kelvin ("absolute zero," 273.16 degrees lower), low-temperature research has brought about such things as household refrigeration and the mass production of pure oxygen and nitrogen from liquid air.

Liquid oxygen represents the nearest large-scale commercial approach to the temperatures of the cryogenic laboratory. Its temperature is not far above the laboratory range. At atmospheric pressure oxygen's boiling point is –182.97 degrees C., only 90.19 degrees above absolute zero.

When the first detectable mists of liquid oxygen were obtained in 1877, it could hardly have been predicted that the product would be distributed in tank-car loads for use in oxyacetylene torches; or that "lox," as the aviators have abbreviated it, would permit the fast burning of fuel in the first long-range rocket, the V-2, and the first supersonic airplane, the X-1. The Linde Air Products Company, pioneer producer of liquid oxygen, began by shipping oxygen as a gas under a pressure of 2,200 pounds per square inch, and still does so to smaller users. For bigger customers, it ships liquid oxygen at –182.97 degrees C. (or –297 degrees on the Fahrenheit scale more familiar to industry). A single tank car carries as much oxygen as 11 freight-car loads of the pressurized gas cylinders.

The big prospect of the air-liquefaction industry lies in the possible use of moderately pure oxygen, or oxygen-enriched air, to speed up combustion in industry, notably in the making of steel. Four fifths of air is inert nitrogen, and this natural mixture can stand artificial improvement for the purposes of combustion. The production of oxygen from the air, of course, leaves four times as much nitrogen as a by-product, and commercial applications for the latter are being sought.

One such application is for assembling products where a tight fit is needed. A common method is expansion fitting, similar to the housewife's trick of putting a tight jar under hot water to expand and loosen the cap. The outside member is expanded by heat and shrinks back tightly on the inside part as it cools. The use of liquid nitrogen, which will quickly chill a piece of metal to –320 degrees F., makes it possible to shrink the inside member first, put it in place, and let it expand to a tight fit as it returns to room temperature. Liquid oxygen has been used for shrinking, but it creates an explosion hazard in the atmosphere. Liquid air has the same disadvantage because the nitrogen boils off first, leaving liquid oxygen again.

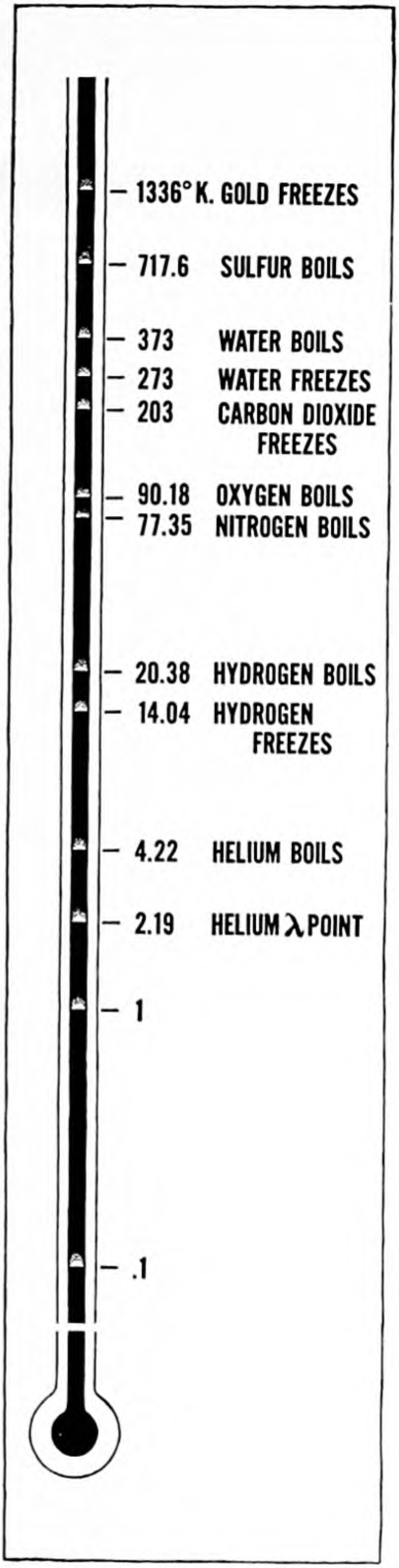
Liquid air, liquid oxygen and liquid nitrogen have thus earned their keep. They are also important as auxiliaries in reaching the temperatures of liquid hydrogen and liquid helium, either to chill them or to interpose an insulating barrier between them and the outside air. In the history of cryogenics, their liquefaction served as milestones along the road toward the still unattained and probably unattainable goal of absolute zero.

The Absolute Scale

Thus far we have used, in about the manner they are normally employed in science and industry, three different scales of temperature and three different definitions of zero. The three scales more or less reflect the history of man's understanding of heat and cold. Gabriel Daniel Fahrenheit, the German physicist who made his living in England and Holland by manufacturing meteorological instruments, set zero at the lowest temperature he could obtain by a freezing mixture. On this scale, "absolute zero" is –459.6 degrees F. In 1742, six years after Fahrenheit's death, the Swedish astronomer Anders Celsius proposed a scale with the freezing point of water at atmospheric pressure as zero and its boiling point as 100. This scale, usually called Centigrade, was officially renamed the Celsius scale last year by the Ninth International Conference on Weights and Measures in Paris. Since the abbreviation C. is retained, the change is not momentous. On the Centigrade or Celsius scale, absolute zero is –273.16 degrees C.

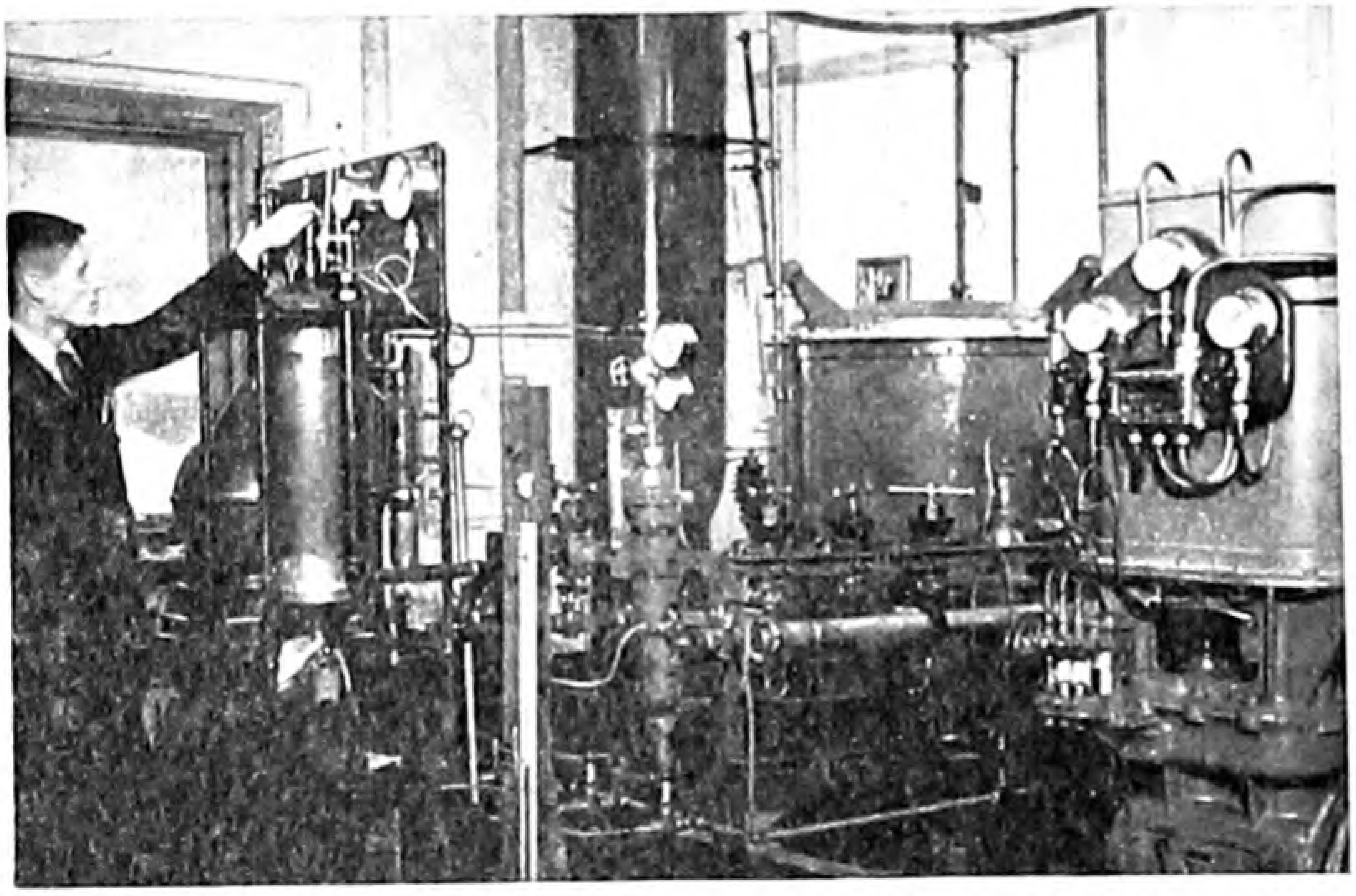
Both the Fahrenheit and Celsius scales are essentially arbitrary, like the pound, the kilogram, the meter or the mile. In 1848, with the increasing thermodynamic sophistication of science, the great Lord Kelvin proposed his "absolute" scale. This retained the Celsius degree but shifted zero to its location in nature. Absolute zero is by definition zero degrees Kelvin, and degrees K. will be used henceforth in this article.

No one has ever reached absolute zero, and it may be stated with a fair degree of confidence that no one ever will. The argument for this is analogous to the indeterminacy principle encountered elsewhere in physics. In order to measure the temperature of a substance, some energy must be exchanged between the substance and its environment. The moment we have energy,



THERMOMETER of significant temperatures in physics is plotted by logarithmic scale to show the detail of the regions close to absolute zero.

THE ENGLISH were among the pioneers of low-temperature physics. This photograph, from David Shoenberg of the Royal Society Mond Laboratory at Cambridge University, shows Sir James Dewar, liquefier of hydrogen.



THE RUSSIANS have been active in the field. The Soviet physicist Peter Kapitza built the first helium-engine type of cryostat. In this picture the engine is at the left. In it the helium cools itself by pushing a piston.

however, we are not at absolute zero. The term implies the absence of thermal energy. Another argument is that the problems of heat removal increase as we get closer to absolute zero. It is not terribly difficult to cool oxygen from its room temperature of about 300 degrees K. to its boiling point of 90 degrees. It is a good deal harder to get to the hydrogen boiling level of 20 degrees. It takes much more effort to attain the liquid-helium level of 4.2 degrees at atmospheric pressure, or 1 degree in a good vacuum. It requires additional magnetic apparatus to chill the helium to .01 or .001 degree. Physicists are still pondering the problem of magnetically lining up the nuclei of atoms to get to the millionth-of-a-degree range. As for closer approaches to the goal of absolute zero, the obstacles appear to be absolutely infinite.

This ultimate barrier does not deter low-temperature investigators any more than the infinite distances of the universe intimidate astronomers. The low-temperature workers have an advantage. From the properties of matter which are dependent on temperature, they know at least where their limit lies.

Descending the Scale

The chronology of man's descent into the depths of temperature goes back to the latter part of the 18th century. The issue then was not low temperature as such, but the question of whether all gases could be liquefied. Some experimenters used pressure, others used cooling, others used both. Ammonia gas was liquefied in 1798. Then, at the Royal Institution in London, Michael Faraday turned his experimental genius to the problem, and in rapid succession liquefied chlorine, carbon dioxide, nitrous oxide and other compounds. By 1835 the French chemist C. Thilorier had produced solid carbon dioxide, the "dry ice" of today.

But try as they might—and they tried mightily by squeezing gases with pressures of hundreds and even thousands of atmospheres—the early experimenters could not liquefy the three most common gases: oxygen, hydrogen and nitrogen. These became known as "permanent gases," apparently resisting the law that any kind of matter could be found in the three states of solid, liquid, and gas. Not until 1869 was it realized that each gas has its critical temperature, above which no amount of pressure will reduce it to a liquid. For the three gases mentioned, the critical temperatures are now known to be:

Oxygen	154.4 degrees K.
Nitrogen	126.1 degrees K.
Hydrogen	33.3 degrees K.

Note that these are the temperatures at which liquefaction will occur under

compression. At atmospheric pressure the corresponding temperatures are lower:

Oxygen	90.1 degrees K.
Nitrogen	77.3 degrees K.
Hydrogen	20.4 degrees K.

To make this list of low-temperature gases complete we may push ahead of our narrative to the discovery of helium in the sun, its extraction from rocks and gas wells on the earth, and its liquefaction at a critical pressure at 5.2 degrees, and at atmospheric pressure at 4.2 degrees.

The liquefaction was achieved in the order of descending temperatures. Oxygen was liquefied in 1877, nitrogen in 1883, hydrogen in 1898. One thing had led to another physically as well as chronologically, for the experimenters generally employed one liquefied gas to cool the next. It was with the aid of liquid air that in 1898 Sir James Dewar brought hydrogen to the temperature from which it could be further cooled to its liquefaction. Thus by the end of the 19th century all the "permanent" gases had been liquefied.

The 20th-century interest in these gases has largely passed from physics to industry. The applications of liquid oxygen and nitrogen have already been discussed. Liquid hydrogen may also have an important future, if experiments at Ohio State University point the way to interplanetary rocket propulsion. A miniature rocket engine has been running there, fueled by liquid hydrogen burned by liquid oxygen. This combination yields the highest exhaust velocity yet attained.

The man who first liquefied hydrogen also invented a device used in every cryogenic laboratory. The Dewar flask, or, to use the common noun, the dewar, is the laboratory counterpart of the Thermos bottle. The double thickness of silvered glass enfolding a vacuum prevents the transfer of heat, thus keeping coffee hot and hydrogen cold. Modern laboratories usually employ a double dewar: a dewar filled with liquid helium immersed in a dewar filled with liquid hydrogen or liquid air. In such an arrangement the barriers to heat conduction, reading from the liquid helium outward, are: glass, silver, vacuum, silver, glass, liquid air or hydrogen, glass, silver, vacuum, silver, glass.

The story of modern research at the lowest possible temperatures is the story of helium, first discovered on earth at the end of the 19th century. At the celebrated cryogenics laboratory of the University of Leiden, run for many years under the benevolent Dutch dictatorship of Heike Kamerlingh Onnes, helium was reduced to the liquid form for the first time on July 10, 1908. It was the last of the gases to yield to man-made cold. By pumping vapor away from the surface of liquid helium, thus cooling the latter by evaporation, Kamerlingh Onnes eventually reached a temperature within .7 degree of absolute zero. This is still about the lowest that can be attained without resort to the newer techniques that will be described presently.

Methods of Chilling

There are a number of ways of achieving low temperatures, and these may be used in various combinations. First there is the refrigeration cycle of the household refrigerator: a vapor is compressed by a pump and cooled by circulating water, to which it yields its heat. This condenses the vapor into a liquid which is allowed to evaporate again, removing heat from the interior of the refrigerator. The method does not suffice to reach truly low temperatures.

One of the classical methods is a simple sequence of compression, heat exchange, and expansion. The gas to be cooled is compressed by a pump, a process which causes it to get warmer. The added heat is removed by passing the gas through a pipe surrounded by another pipe containing a colder liquid. In making liquid air, for example, the cooling may be done by water; in making liquid hydrogen, the cooling may be done by liquid air; and in making liquid helium, the cooling may be done by liquid hydrogen. The compressed and cooled gas is now allowed to expand through a narrow orifice, which cools it even more.

This cooling by expansion was first proposed by the British physicists James Prescott Joule and William Thomson, later Baron Kelvin of the degree K., and for them it is named the Joule-Thomson effect. Often the gas cooled by the foregoing process is used to cool the incoming gas, so as the process goes along the gas gets colder and colder until it finally liquefies. In the case of helium the Joule-Thomson effect operates only after the gas has been cooled to the temperature of liquid hydrogen.

Another method, theoretically obvious long ago but only recently applied on a major scale, is to allow helium to drive an engine so that it gives up its thermal energy in mechanical motion. Helium at room temperature is something like steam far above the boiling point of water. Under the right conditions it will drive a small version of a steam engine, and just as the spent steam turns to water the exhaust helium turns to liquid. All this is easier said than done because no lubricant will serve to ease friction at the temperature of liquefying helium, and the engine must be built with such close mechanical tolerances that only a thin stream of helium gas will escape between the piston rings and the cylinder wall.

Peter Kapitza, the Russian physicist who is now presumably engaged in atomic research, developed such a helium engine while working in England before the war. He chilled helium to the temperature of liquid air and ran it through a one-cylinder engine from which it emerged as a liquid. C. T. Lane of Yale University built a similar apparatus, thus establishing one of the first few laboratories in the Western Hemisphere capable of working with liquid helium. E. F. Burton of the University of Toronto and W. F. Giauque of the University of California had earlier used other methods of attaining such temperatures.

Until only a few years ago every low-temperature laboratory had to build its own low-temperature apparatus. The research worker had to be a first-class refrigeration engineer, one result being that every cryogenic apparatus was unique. Another result was that there were few cryogenic laboratories. The situation has changed with the development of a helium-engine type of apparatus by S. C. Collins of the Massachusetts Institute of Technology. It is a two-cylinder engine, with the cylinders arranged so that the cold exhaust gas of one cools the intake gas of the other. With this apparatus, and without the assistance of intermediate coolants such as liquid hydrogen or liquid air, it is possible to proceed directly from the temperature of cold tap water to that of liquid helium. The helium is compressed in powerful compressors, cooled by water and sent through the two-stage engine.

An ex-student and co-worker of Collins, D. O. McMahon, left M.I.T. to join the laboratories of Arthur D. Little, Inc., a few blocks up Memorial Drive in Cambridge. There, in addition to carrying on low-temperature research of his own, McMahon has put the Collins Helium Cryostat into what, for this field, can be regarded as mass production. Nineteen of the machines have already been shipped to laboratories all over the country. The Naval Research Laboratory alone has three. This means that a physics laboratory, having decided to go into low-temperature research, does not have to wait a year or two until its physicists build a cryogenic apparatus and perfect it. At a cost of about $22,000, the laboratory can be in business almost immediately. Collins is now building a gigantic cryogenic apparatus for himself, about which workers in other laboratories speak with awe. The rumor is that it will provide a working space, all at the temperature of liquid helium, as big as a large refrigerator.

The Properties of Helium

Helium might be called the "less" gas. It is colorless, odorless, tasteless, and so nearly weightless that its principal use is for the inflation of balloons. Hydrogen

is even lighter, but hydrogen burns, as it did in the Graf Zeppelin. Helium is one of the "noble" gases; it disdains to react with other elements. It was first discovered in the spectrum of the sun, where it is believed to be the final inert product of the nuclear reaction which is the source of the sun's radiant energy. On earth it is created by the radioactive breakdown of heavy elements such as uranium and radium. Their alpha rays are the nuclei of helium atoms, consisting of a family of two protons and two neutrons. These particles capture a pair of electrons as satellites and become helium atoms. Nuclear physicists sometimes reverse the process, stripping away the electrons to use the alpha particles as projectiles in such machines as the cyclotron.

On the current frontier of cryogenic research, helium plays a dual role. As a liquid, it is a curious and fascinating substance that is occupying the full attention of many experimenters and theorists. As a cold liquid bath, it causes substances immersed in it to exhibit the curious and fascinating properties of all matter shorn of most of its thermal chaos.

The fascination of liquid helium itself derives from phenomena that occur when its temperature is reduced to 2.19 degrees K. Picture a dewar of freshly made liquid helium inside a dewar of liquid air. There are narrow vertical slits of unsilvered glass on both dewars, and when they are turned to coincide, a window is formed through which the experimenter observes the liquid. Both the liquid air and the liquid helium, frigid though they be, are boiling just as water boils in a teakettle at 373.1 degrees K. The air bubbles are large; the helium bubbles are small. The outside dewar of liquid air is vented, keeping it safely at atmospheric pressure. The inside dewar of liquid helium is connected to a vacuum line which draws off the vapor so that the remaining liquid will get still colder.

As the helium reaches 2.19 degrees, the boiling seems suddenly to stop. The surface is as smooth as glass, although jellylike ripples may run across it. Its temperature continues to drop. Vapor is still formed and drawn away, but it escapes without noticeably ruffling the surface. The helium has evidently entered a new state. Because of many other curious properties, it has been given the name helium II.

Is it still a liquid? Liquids have viscosity, and in some respects helium II has none. It flows through the narrowest of orifices. Low-temperature workers were annoyed by the difficulty of making leakproof containers for it until they realized that this was demonstrating a new kind of matter.

Is it a gas? Helium II atoms are more mobile than the atoms and molecules of gases. But they obey the law of gravity as a liquid does, remaining at the bottom of containers. They also form a surface which will seek its own level more diligently than water.

Is it a solid? One would expect that if a substance liquefies at 4.2 degrees and undergoes another change of state at 2.19 degrees, the second change would be freezing. In fact, when this state was discovered there were attempts to show that it was some kind of crystalline array. There is no solid as agile as helium II; it flows, pours or dances through the narrowest crevice. Besides, there is a solid helium. While it cannot be obtained under atmospheric pressure with the lowest temperatures yet reached, it has been obtained by building up the pressure to 25 times that of the atmosphere.

Physicists have worked out a diagram of the states of cold helium that may be seen at the top of page 36. Because the intersection of two of the lines on the graph is reminiscent of the shape of a slightly tilted Greek letter λ, or lambda, cryogenics has adopted the term "lambda point" for the transition to helium II.

If the journey past the lambda point does not take us to a solid, a liquid, or a gas, what is helium II? Physicists have been driven to calling it "the fourth state of matter." Some of them describe it as "the quantum fluid," of which more later. The phenomena that urgently require explanation are "the creeping film," "the fountain effect" and "second sound."

Let us begin with the case of the creeping film. If the bottom of a small vessel is placed in helium II, an unlikely thing happens. The liquid helium climbs up the sides of the vessel and fills it to the same level as the surrounding surface. If the vessel is then lifted, the helium climbs back over the edge and down to the surface again. Water or any other well-behaved liquid will, of course, find its level but such liquids need a pipe or a primed siphon. In some peculiar way helium II gets around on its own.

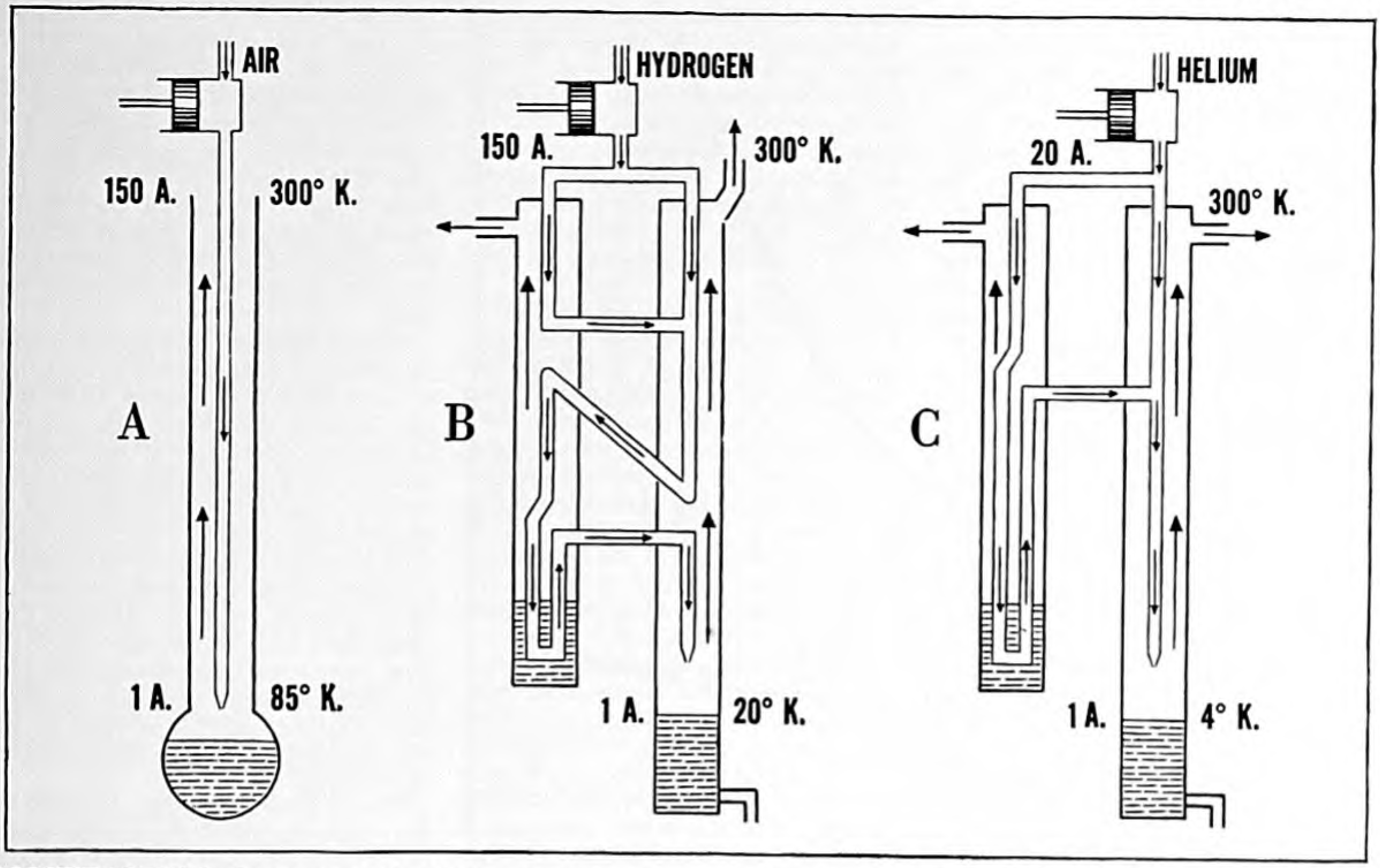
The fountain effect occurs when a vessel which narrows to fine tubes at top and bottom is lowered into helium II. Sometimes the bottom of the tube is filled with finely ground powder. Helium II infiltrates the lower tube with ease. If the vessel is then warmed by light, helium will spurt through the upper tube in a spectacular fountain.

Second sound is related to the fact that helium II is a superlative conductor of heat. Warm one end of a vessel, and a pulse of heat will be quickly conveyed to the other end. This phenomenon might be described as a heat wave propagated quickly through the coldest kind of matter. Because it is analogous to the pressure waves of sound, the Russian physicists who first discovered the phenomenon in Moscow near the end of the war called it second sound. Its existence, which had been predicted by both Russian and American physicists, was soon confirmed by Lane at Yale. The phenomenon is now being studied intensively in many laboratories. Pellam, for example, sends pulses of heat through helium II at different temperatures, and by electronic methods has displayed the lapse of time between the transmitted and received temperature signal on a cathode-ray tube.

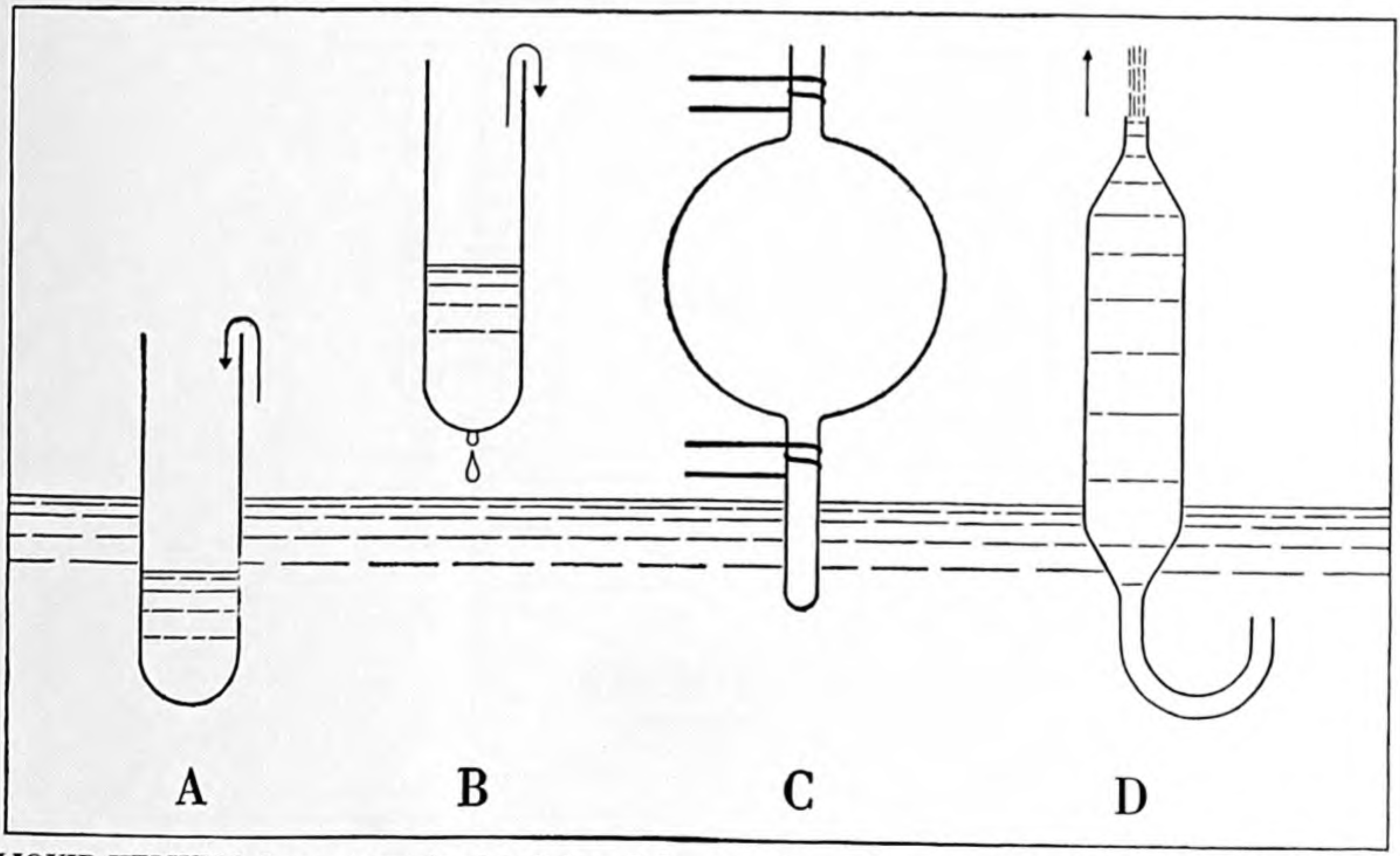
In cryogenics, unlike other branches of physics, there appears to be a free exchange of information among nations. U.S. low-temperature workers keep a close watch on the Russian journals. A typical U.S. paper in low-temperature physics will bear footnote references to "J. Phys. U.S.S.R.," citing the theories and experiments of Kapitza, Landau, Peshkov and Andronikasvilli. In this coldest of all scientific disciplines, the cold war is directed against the secrets of nature.

One of the most revealing eccentricities of helium II was predicted by Landau and in 1946 was observed by Andronikasvilli. His experiment was to rotate a vessel of chilled helium and to measure its inertia. As the temperature dropped below the lambda point, the vessel's resistance to acceleration rapidly decreased. The only explanation was that a considerable number of the atoms in the helium II were not participating in the rotation. This component, as Collins puts it in a recent survey, "does not take part in the rotary motion, but glides through the interpenetrating atmosphere of normal helium atoms without friction." It is as though a man in a tightly packed crowd could remain motionless while the crowd surged past him.

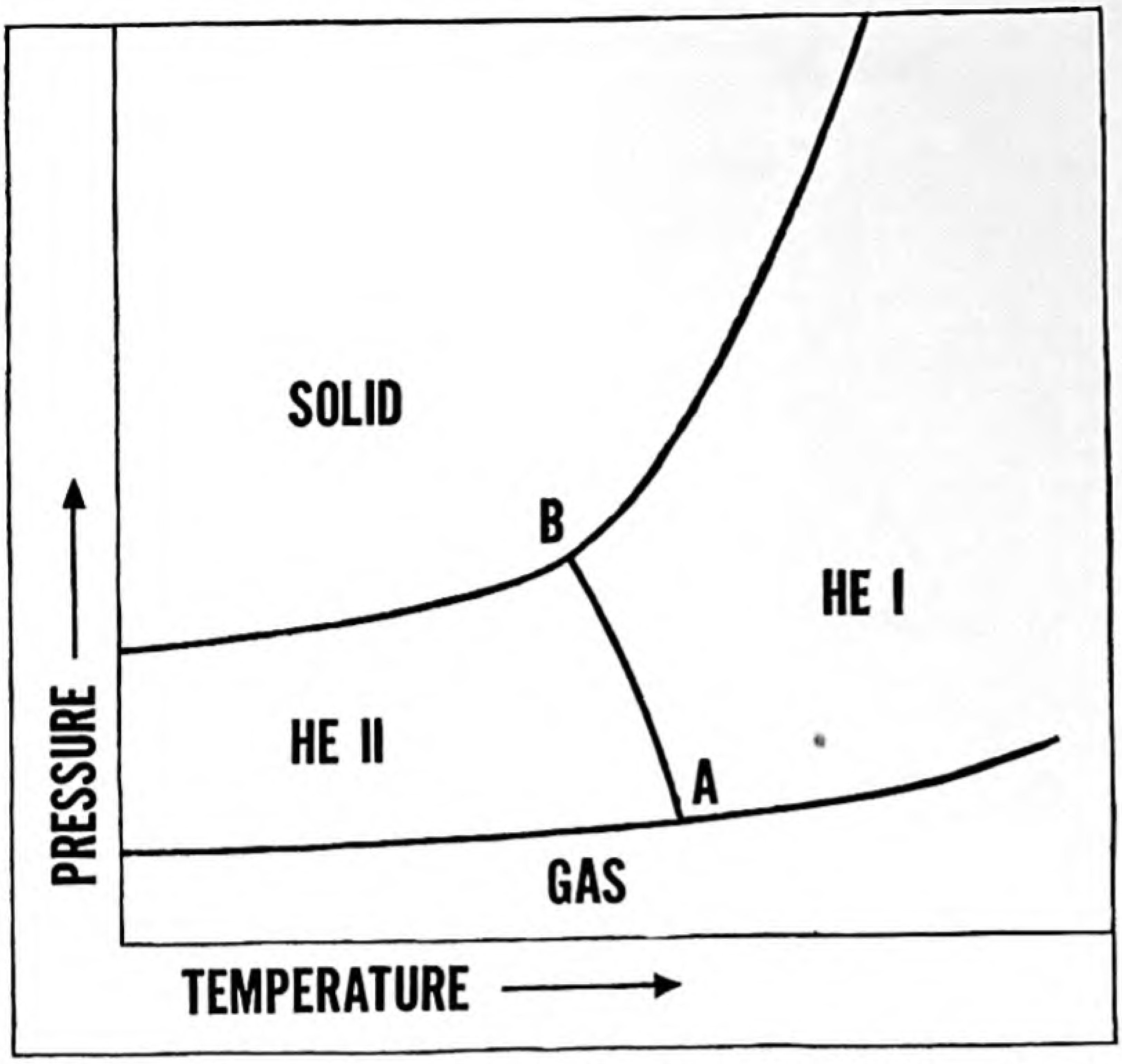
What is the explanation for these strange occurrences in helium II? Physicists have adopted the concept that, even at a degree or two above absolute zero, an increasing proportion of the helium atoms drop to a "zero energy state." It is here that the concept of the quantum fluid is necessary. The quantum theory states that the energy of an atom cannot be gained or lost continuously but must come in certain prescribed amounts. A graph showing the loss of energy with temperature would be a series of steps rather than a smooth curve. At ordinary temperatures the steps are so small in proportion to the total thermal energy that the individual atomic quanta do not mar the smoothness of the curve. At a degree or so above absolute zero, however, we are on the last few steps from the bottom of the staircase. If more energy is withdrawn, it is no longer possible for all of the atoms to shift to lower levels while maintaining a normal "probability distribution." Some of the atoms on the step next to the bottom will lose all their remaining thermal energy with the emission of a single quantum. They will drop to the



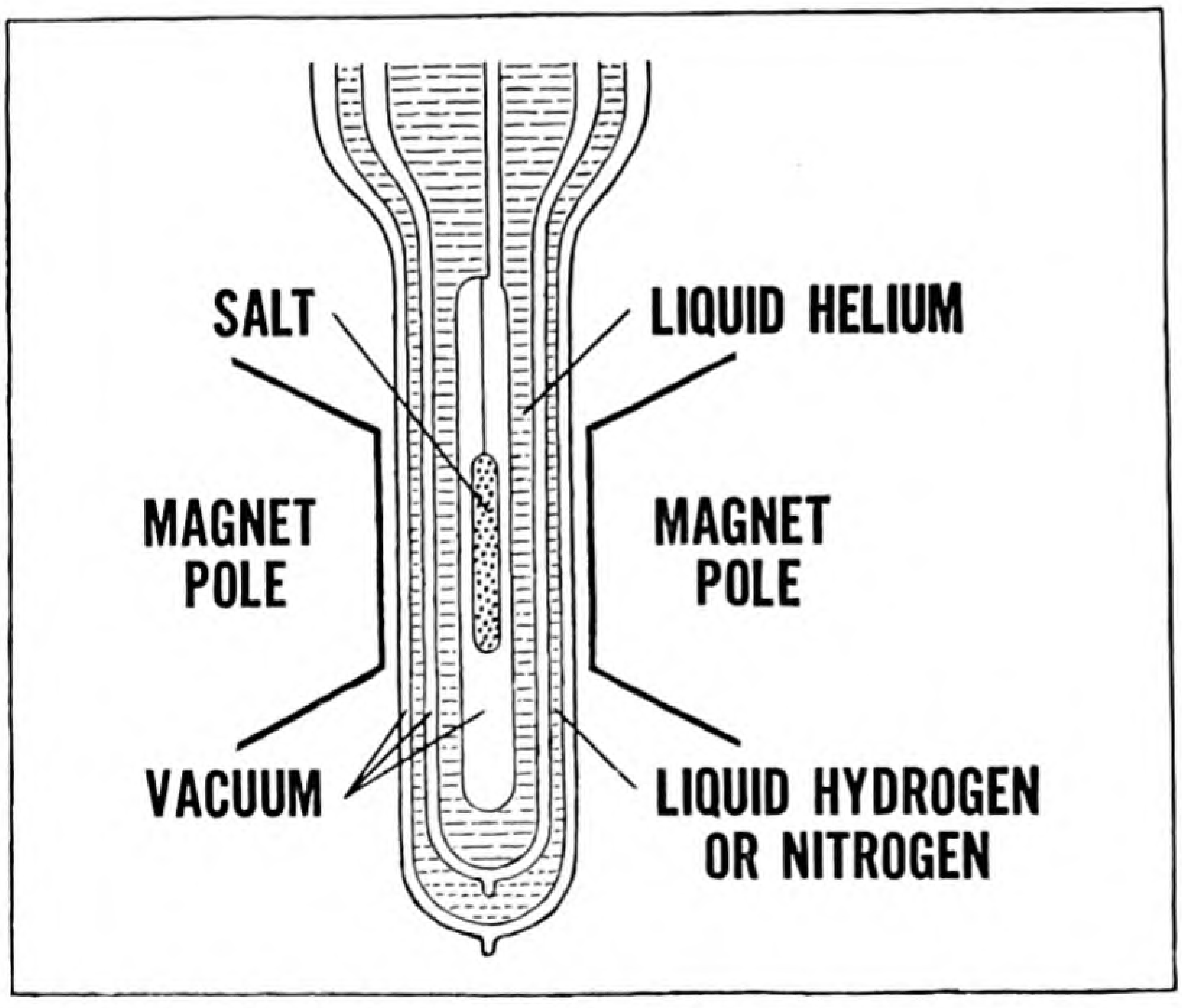
LIQUEFACTION of air was accomplished in early apparatus (A) by compressing it to 150 atmospheres, allowing it to expand through a nozzle, and circulating the gas thus cooled around the incoming gas. The latter was then made progressively colder until it liquefied. Lower temperatures of liquid hydrogen and helium were attained by the same means (B and C) except that gas was circulated through another gas already liquefied.



LIQUID HELIUM behaves strangely when it is cooled to 2.19 degrees K. If a vessel is lowered into it (A), the helium climbs in. If the vessel is lifted out of it (B), the helium climbs out. The speed with which this film travels may be measured by timing its arrival at the top wire on a glass bulb (C). If helium is allowed to enter a fine tube at the bottom of another vessel (D), it will spurt from the tube at the top when light is shined on it.



PHASE DIAGRAM of helium shows the relationship of helium II to the usual three phases of an element: gas, liquid and solid. The transition from helium I to helium II is called the lambda point because the intersection of line AB with others is roughly comparable to the Greek letter λ, or lambda.



LOWEST TEMPERATURES are attained by suspending a certain salt in a container of liquid helium between the poles of a magnet. Helium gas is first allowed into vacuum chamber to cool salt. Gas is removed and salt is further cooled when magnet lines up molecules and stills thermal motion.

"ground state," while the others remain one, two, or more steps up.

These ground-state atoms are presumably responsible for the many foibles of helium II. The colder the helium, the more of its atoms will be in the ground state. Thus there is a mixture of two fluids, a normal fluid and a "superfluid," which do not interact with each other except that with the loss or gain of heat an atom can go from one state to the other.

The ground-state or superfluid atoms are believed to remain aloof when a vessel of helium II is rotated. It is they that find their way through the narrowest crevices, that creep up the walls of containers. The fountain effect may occur on a one-way street policed by a kind of quantum traffic cop. The ground-state atoms proceed through the narrow channel, and at the far end are warmed by an outside source of energy. This raises them to a normal state in which they cannot get back through the narrow entrance. Meanwhile more ground-state atoms are coming in, with the result that enough pressure might be built up to create a fountain.

Second sound is likewise explained by the transformation of ground-state atoms outward from the source of heat in a sort of wave front. The ground-state atoms move back to replace those that have been heated. Then they slide through the normal atoms without friction.

These are the somewhat crude visualizations of a theory based on quantum-mechanical mathematics. The theory was derived by Laszlo Tisza of M.I.T. and Fritz London of Duke University from a mathematical generalization of Albert Einstein and the Indian physicist S. N. Bose. The "Bose-Einstein statistics," the explanation of which is beyond the scope of this article, were expected to apply to the abundant isotope of helium which has an atomic weight of 4, but not to the rare isotope which has an atomic weight of 3. The theorists predicted that if a supply of He^3 became available, it would not enter the quantum-fluid condition at low temperature. In other words, experimentalists could expect no He^3 II.

Several groups have therefore attempted to achieve the separation of helium 3 by means of the fountain effect and similar processes depending on superfluidity. At Ohio State and Yale a concentration of helium 3 was obtained by utilizing the fact that it did not act as a superfluid below the lambda-point temperature for ordinary helium.

Meanwhile the powerful devices at the disposal of the Atomic Energy Commission were put to work on the problem. In the nuclear reactor, hydrogen 3, or tritium, was being made by transmutation. As it decayed by radioactivity it turned into helium 3. With this first pure supply of the rare helium isotope,

the customary tests for a superfluid were re-enacted.

Before revealing the denouement to those readers who did not read it in *The Physical Review* or in the reports which appeared in some newspapers and magazines, the suspense may be heightened by pointing out that this was to be the crucial test of the London-Tisza, or "American," theory of superfluidity as opposed to that of Landau in Moscow. Both theories had accounted in different ways for phenomena observed up to that time, and both had scored successful predictions of earlier experimental results.

As recently as August of last year Tisza, writing in *Physics Today*, said: "If the Bose-Einstein statistics is essential for the superfluidity of the abundant isotope He^4, then He^3 should be not superfluid. . . . On the other hand, according to the other theories, no essential difference can be expected for the two isotopes."

Tisza concluded: "If the separation of He^3 in sizable amounts is indeed possible, the study of this substance would be of considerable interest. It would be the only case where two stable isotopes have radically different properties. It seems very likely that He^3 cannot exist in the liquid state at all. Such a liquid should have a vanishing dynamic viscosity and a high kinetic viscosity. Either we will have a liquid of entirely unheard-of properties, or the system will avoid the dilemma of the large and small viscosities by not liquefying at all, but will either freeze or rather stay a gas at vanishing pressure and temperature. It is to be hoped that the experimental decision of this question will be forthcoming before long."

The experimental decision came from the laboratories of the AEC at Los Alamos in January, 1949. S. G. Sydoriak, E. R. Grilly and E. F. Hammel showed that helium 3 does become a liquid, at 3.2 degrees. In March, however, D. W. Osborne, B. Weinstock and B. M. Abraham of the Argonne National Laboratory announced that helium 3 does not pass over to a superfluid in the same manner as helium 4, even when it is cooled to 1.05 degrees. Thus the principal London-Tisza prediction of the differences between the two isotopes was borne out. Tisza's further prediction that helium 3 would not liquefy at all was obviously not. Whether liquid helium 3 will have "entirely unheard-of properties," opening a new vista of research as absorbing as that of He^4 II, remains to be seen. The wide range of possibilities that Tisza permitted himself to suggest in advance of the experiment is a good index of the superfluid state of low-temperature physics. Almost anything can happen when helium 3 is made still colder. Workers outside the AEC are eagerly waiting their chance to try it, meanwhile exploiting the proved difference between the two isotopes to do some more efficient concentration by the thermal methods available to them. Meanwhile the astonishing disparity between two substances which differ only by a single neutron in the nucleus will give the nuclear physicists something to ponder.

Superconductivity

Perhaps the most impressive of all phenomena at temperatures within a few degrees of absolute zero, and certainly the most intriguing from the practical engineer's point of view, is the conversion of certain metals into superconductors. At a particular temperature, which is different for each of the metals involved, a wire will lose all measurable resistance to the flow of electric current, and a sheet or disc will become an efficient screen against magnetism.

The fact that electrical resistance drops with falling temperature is no surprise. The resistance of a wire is generally proportional to its temperature; the filament of an electric light bulb will allow more current to enter when it is first switched on than after it becomes incandescent. This effect is put to work in electric-resistance thermometers, which have been standardized for both high and low temperatures. On this principle, the resistance of a wire should fall to zero at absolute zero, and the rate of decrease should be gradual as the temperature drops. This is what happens with the best of normal-temperature conductors: copper and silver.

In the case of many other metals, including some which are poor conductors at ordinary temperatures, resistance vanishes completely at a transition point from a fraction of a degree to several degrees above absolute zero. In these materials, which include mercury, tin and lead, a current once started will continue to flow practically forever unless something is done to destroy the superconducting condition.

The phenomenon is not entirely new. Kamerlingh Onnes discovered it in 1911 when he attempted to explore the lower end of the temperature-resistance curve. He tested solidified mercury in a bath of liquefied helium. Down to 4.3 degrees the mercury's resistance steadily decreased, as expected, to about one five-hundredth its value at the freezing point of water. Then it suddenly dropped to less than a millionth of the normal value. Further research has shown that the purer the metal, the more completely and quickly does its resistance vanish.

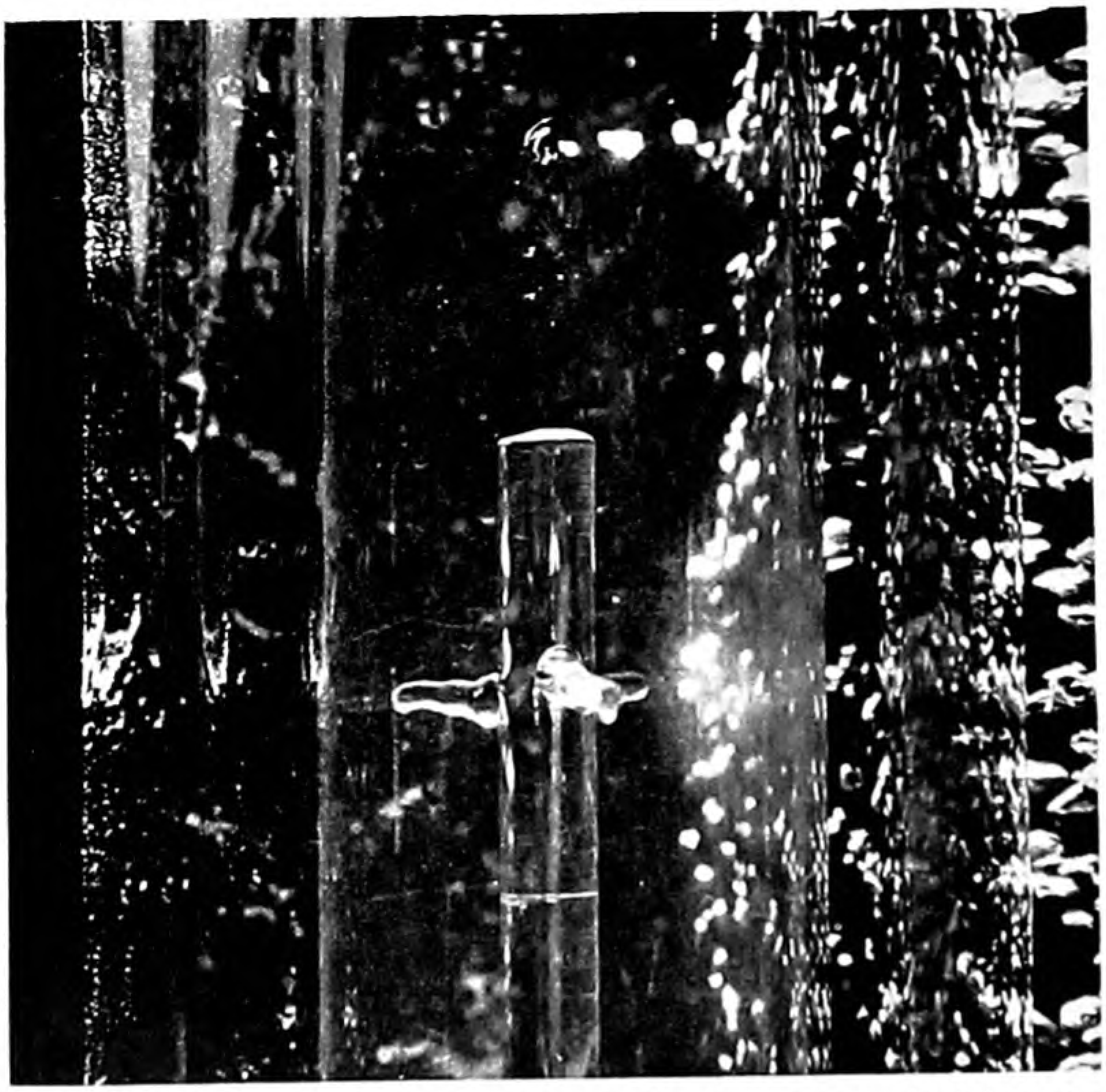
The metallic superconductors since identified, in addition to mercury, tin and lead, include aluminum, zinc, thallium, indium, tantalum, gallium, thorium, titanium, columbium, vanadium, cadmium, zirconium, hafnium and lanthanum. Columbium, it was found in 1930, reaches superconductivity at the relatively high temperature of 9.22 above absolute zero.

Naturally such discoveries stirred the hope that superconductivity might be applied to reduce the huge problems of electric-power distribution which may be traced simply to electrical resistance. Efforts were made to find alloys and compounds in which the transition point would be nearer to practical temperatures. It was disappointing that most alloys showed a transition temperature even lower than those of their constituent elements. The greatest advance in transition temperature was scored with nitrides, carbides and borides of metals. The most successful of these is columbium nitride, which becomes a superconductor at around 15 degrees. This temperature does not necessarily require the use of liquid helium; it can be obtained with hydrogen boiling under reduced pressure.

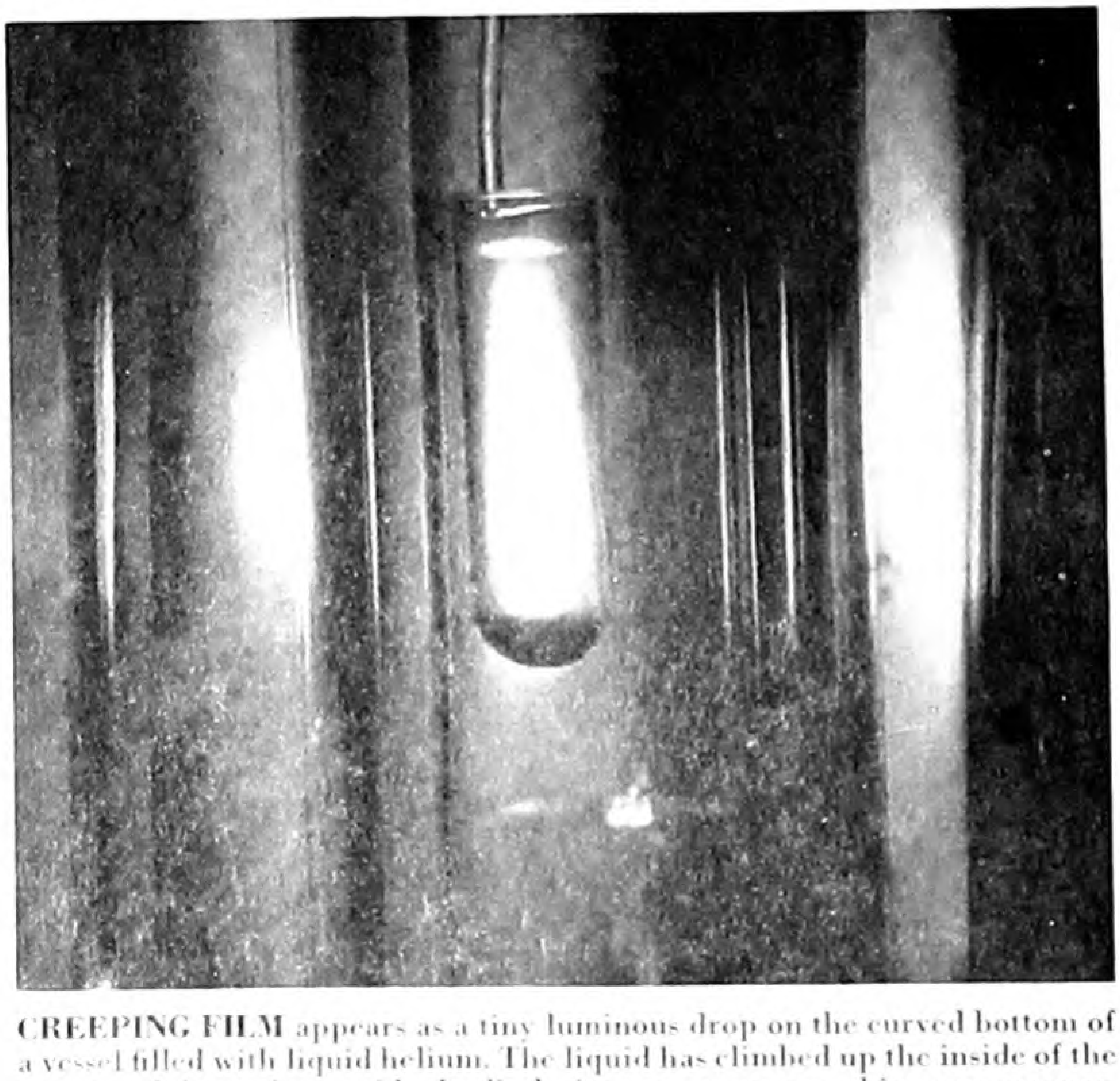
Columbium nitride is being studied intensively at Columbia University and at Johns Hopkins University. During the war Donald H. Andrews of Johns Hopkins cleverly applied the properties of the transition state to make an extremely sensitive detector of infrared radiation. In a substance just on the borderline between the superconducting and non-superconducting condition, a slight change of temperature will cause an enormous change in electrical resistance. The on-the-verge material is also extremely sensitive to other outside influences of an electromagnetic nature, as was discovered when the Johns Hopkins superconducting bolometer began pouring music from a nearby radio station through its amplification system. The columbium nitride had acted as a detector of radio waves. It will also detect alpha rays, and steps have been taken to develop an extremely sensitive alpha detector.

The ability to detect or rectify an alternating current at low temperatures is not confined to columbium nitride. F. G. Dunnington and Bernard Serin of Rutgers University have subjected a helium-chilled tin wire to a combination of direct and alternating current which sweeps it in and out of the superconducting state as often as the alternating current reverses itself. At Yale, Amherst College and elsewhere, currents of assorted frequencies, including the microwaves that have been inherited from radar, are being employed to explore the mechanism of superconductivity.

The key to the situation appears to be magnetism. Probably the most rigorous definition of a superconductor is not its lack of resistance, but the fact that it is "a perfect diamagnetic." This means that, in direct contrast to a piece of soft iron, it rejects any attempt to impose a magnetic field upon it. Lines of force emanating from a nearby mag-



FOUNTAIN EFFECT of liquid helium is photographed in the laboratory of Arthur D. Little, Inc. Glass tube in center is inside Dewar flask containing liquid helium. Top of fountain is just visible near the top of picture.



CREEPING FILM appears as a tiny luminous drop on the curved bottom of a vessel filled with liquid helium. The liquid has climbed up the inside of the vessel and down the outside. It climbs into an empty vessel in same manner.

net or coil, which would be drawn in by iron, are pushed aside by a superconductor and made to detour around it.

Perhaps the most dramatic demonstration of this property is the floating magnet. If a permanent magnet is dropped into a dewar containing a lead plate immersed in liquid helium, it will not fall to the plate. It remains suspended some distance above the superconducting lead. Occasionally it darts from side to side, but it does not come down.

The invisible strings controlling this strange marionette show are magnetic lines of force around the superconducting metal in the helium bath. The approach of the magnet generates electric currents in the surface of the lead plate. The magnetic effect of these currents precisely neutralizes and opposes that of the magnet, which has made a mirror image of itself in the surface of the superconductor.

Such curious powers are limited. If the field of the magnet is strong enough, it will penetrate the chilled metal. This destroys superconductivity. The more the temperature drops below the transition point, however, the more magnetic force it takes to destroy superconductivity. Another way of disrupting the superconductive state is to send a sufficiently strong current through the metal. This phenomenon is now interpreted as the result of the current's magnetic field. The lower the temperature in relation to its transition point, the more current the superconductor will be able to superconduct.

The most interesting demonstration of these characteristics occurs right at or just below the transition point. Here an intermediate state exists between superconduction and normal resistance, in which the balance can be swung one way or the other by small amounts of current. It is here that some of the effects described earlier are achieved: the conversion of alternating current to direct, the detection of alpha radiation from a radioactive atom, of infrared radiation from a warm body, and of long- or short-wave electromagnetic radiation from a radio station.

What is the future of this department of low-temperature physics? Its workers believe they can learn more about the essential nature of electricity and magnetism, the behavior of which is in some ways simplified by the suppression of random thermal noise. They are already beginning to examine some long-established notions in a different light. The surprising thing, says one cryogenics investigator, is not that metals at low temperatures superconduct, but that those at ordinary temperatures do not.

Air is a superconductor of radio waves; glass is a superconductor of light; helium II is a superconductor of heat. Copper wires at ordinary temperatures, though they may be regarded as good conduc-

tors by the engineer, are in a sense opaque to electrons. The flow of current is impeded by friction, a good thing for lighting an incandescent lamp, but a bad thing for getting the current from the power plant to the lamp. The trouble is that the moving electrons interact with their medium. When certain metals become sufficiently cold, the electrons stop interacting and become as free as the ground-state atoms of helium II. To them the metal becomes transparent.

Another view of superconduction is to think of every atom, in its own private sphere, as a superconductor. Electrons move in their orbits with no friction, much as the planets circle the sun with no resistance in the vacuum of space. In its own atomic orbit, there is nothing to stand in the electron's way. A superconducting piece of metal can be considered the joining of countless atoms in such a manner that each electron finds a superorbit, with the rules of quantum mechanics barring any other particle from trespassing upon its right of way.

The Lowest Temperatures

While scores of physicists are exploring superconductivity and helium II, extending the work of those who attained the temperatures of liquid helium, others are venturing into the remote region between zero and 1 degree K. In the approach to absolute zero, nature's opposition is fierce. A formidable technique must be brought into play: adiabatic demagnetization.

This remarkable procedure was suggested independently in 1926 by Peter J. W. Debye, now head of the Cornell University chemistry department and no longer a worker in cryogenics, and W. F. Giauque, whose elegantly equipped laboratory at the University of California was the headquarters of the semiannual conference of ONR cryogenics contractors in February. Giauque and W. J. de Haas of Leiden perfected the method, and now most of the cryogenic laboratories are already using or building or planning the magnetic apparatus for getting below 1 degree K.

The method is skillfully based upon a special kind of thermal chaos: the random arrangement of the magnetic poles of individual molecules in substances known as paramagnetic salts. The molecules of most substances are magnetically neutral. In those we are about to consider, such as iron ammonium alum, each molecule has one spinning electron which is not magnetically canceled out by other electrons in the same molecule. Under ordinary conditions, these molecular magnets are distributed in such a way that, on the large scale, they do cancel each other out. A handful of the salt will thus exhibit no inherent magnetism to an outside observer. If an outside magnetic field is applied, however, the molecules will tend to line up like tiny compass needles. The regimentation, being the antithesis of thermal chaos, squeezes some of the heat content out of the salt.

This may seem abstruse and remote, but the laboratory practice can be reduced to a procedure not unlike that outlined in a cookbook. Put a pinch of the salt, or as much as a pound of it, in the bottom of a dewar. Pour liquid helium over it, perhaps a quart per pound. Allow to cool by boiling the helium at lower and lower pressure, removing its vapor with a vacuum pump. Observe the lambda-point transition to helium II. Keep the vacuum pump going until the temperature drops to 1 degree. Now apply a strong magnetic field. Allow time for the molecules to become fixed along the direction of the lines of force, and for the helium II to remove the heat. Remove the magnetic field. Observe the drop in temperature. Repeat.

The drop in temperature occurs in the final step of each such process. While in the regimented magnetic state, the salt has given out heat that the helium carries away. When the magnetic field is removed (usually by swinging the dewar away from the fixed pole pieces), the molecules "relax," turning in all directions. This brings a concomitant drop in temperature. The salt needs heat energy to become disorderly, and this it absorbs from the helium; thus the helium-salt combination cools off during demagnetization.

Depending on the strength of the magnetic field, the temperature can be reduced to .1 degree, .01 degree and even .001 degree. Thus powerful magnets have been brought into cryogenic laboratories, just as they have been brought into the laboratories of nuclear physics. The Naval Research Laboratory at this writing is installing a powerful electromagnet with a field strength of 100,000 gauss, known as the Bitter magnet because it was originally designed for spectroscopic work by Francis Bitter of M.I.T. The turns of the magnetizing solenoid are thick copper straps which carry a current of thousands of amperes, are supplied with power by a 2,000-kilowatt generator, and must be cooled with a flow of 800 gallons of water per minute. If the cooling-water supply should fail, the copper would start boiling in a few seconds. Warren E. Henry, a physicist trained at Tuskegee Institute and the University of Chicago and an alumnus of Collins' laboratory at M.I.T., expects to cool as much as a quart of helium to .01 degrees absolute when the Naval Research Laboratory apparatus gets going on a full scale.

One of the things that happens at these temperatures is the formation, above the helium surface, of the world's most perfect vacuum. The vapor pressure of the helium drops so low that there is as much chance of an atom escaping into the space above the surface as there would be of a piece of iron boiling out of the structure of a building at room temperature. Air or anything else that possibly leaked into the dewar would freeze solid. In the space just above helium close to absolute zero, there is just about absolute emptiness.

In this cold, still realm, physicists are checking the basic concepts of thermodynamics that were set forth in the days when physics could only deal with the thermal chaos. They are giving a more direct physical meaning, in terms of atoms and their electrons and nuclei, to such statistically derived words as heat, entropy and temperature. They may emerge with a new definition and a more accurate localization of "absolute zero."

Already there are some who are not satisfied with the prospect of reaching .001 degrees K., which is the most to be expected from regimenting and relaxing the spin of electrons. They are looking ahead to using the last conceivable resource: the magnetism of atomic nuclei. This, according to the theoreticians, offers the possibility of temperatures that will not exceed absolute zero by more than a millionth of a degree.

To this writer, the laboratory that best represents the varied prospects, abstruse and practical, of cryogenic research, is that of Ohio State University. For the attainment of extremely low temperatures by electron demagnetization, a current-carrying coil is used to set up a magnetic field of 4,000 gauss. The coil is cooled by liquid nitrogen. This does not attain superconductivity, for which liquid-helium temperatures would be needed, but it does exploit the normal drop of electrical resistance with reduced temperature. Thus the lowest temperatures of an earlier era of cryogenic research abet the attainment of new depths; cryogenics lowers itself by its bootstraps. When workers in the field are asked if there is any possibility of ever using real superconductivity for the powerful electromagnets of cryogenics and nuclear physics, they will admit that it does not seem practical. But after a few moments meditation they are likely to remark: "Well, at Ohio State, you know, they're using liquid nitrogen. . . ."

The Author

HARRY M. DAVIS was the author of "Radio Waves and Matter" and "Mathematical Machines," which appeared in past issues of SCIENTIFIC AMERICAN. A week after he had completed this article, Davis was drowned in a swimming accident at Biloxi, Miss. He was one of the rare professional journalists whose reporting of science was admired by scientists. His death was a grievous loss to the journalism of science.

Bibliography

PHENOMENA AT THE TEMPERATURE OF LIQUID HELIUM. E. F. Burton, H. Grayson Smith and J. O. Wilhelm. Reinhold Publishing Corporation, 1940.

SCIENTIFIC AMERICAN October 1956, Vol. 195, No. 4, pp. 93-102 OFFPRINT 207

MESONIC ATOMS

by Sergio De Benedetti

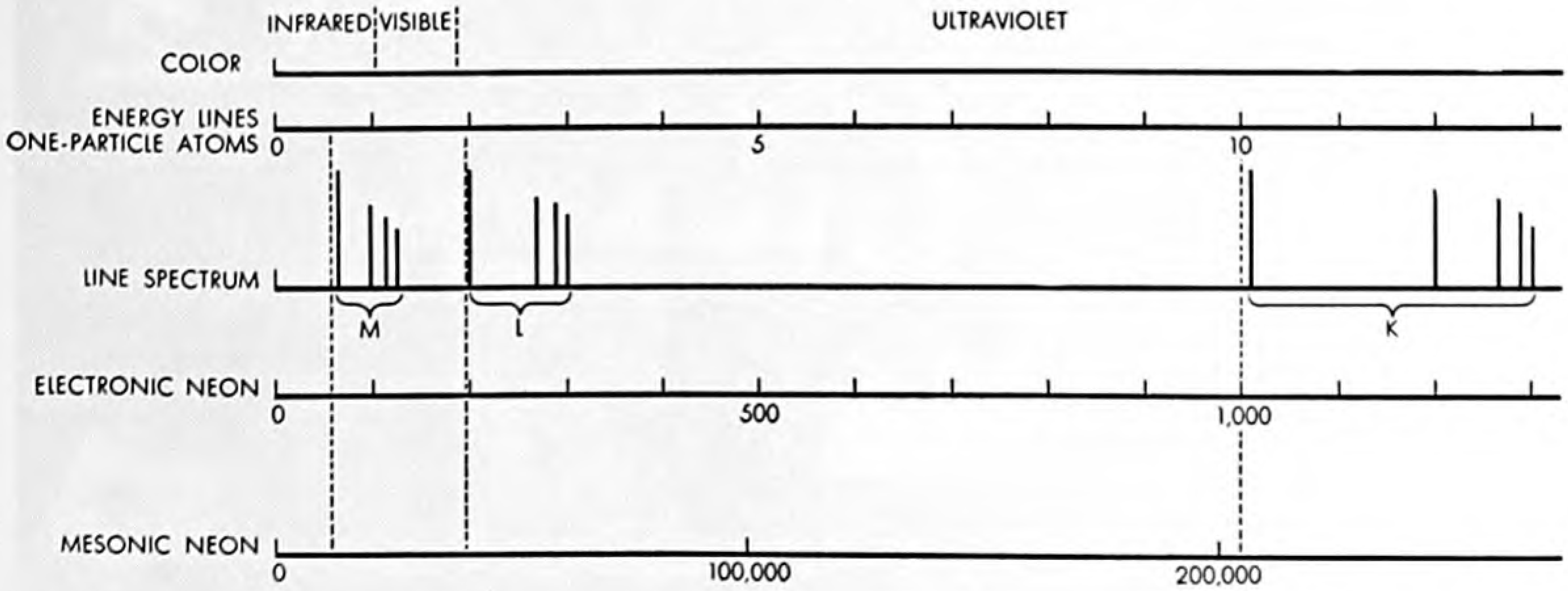
For brief instants mesonic atoms may spin like electrons on orbits around atomic nuclei. The X-rays they emit, as they jump from orbit to orbit, may illuminate the nature of the nuclear binding forces.

In the half century since the atom was opened up as a new world to explore, physicists have been busy taking it apart to see what it is made of. It has been a period of violent bombardment and assault, and experimental work in atomic physics has become popularly known as "atom smashing." But the study of the atom is now entering a new phase. Nowadays atoms are so well understood that physicists can undertake to build as well as destroy them. Completely artificial atoms have been forged from some of the newly discovered atomic particles, and these serve as tools for testing theories about the nature of the atomic world.

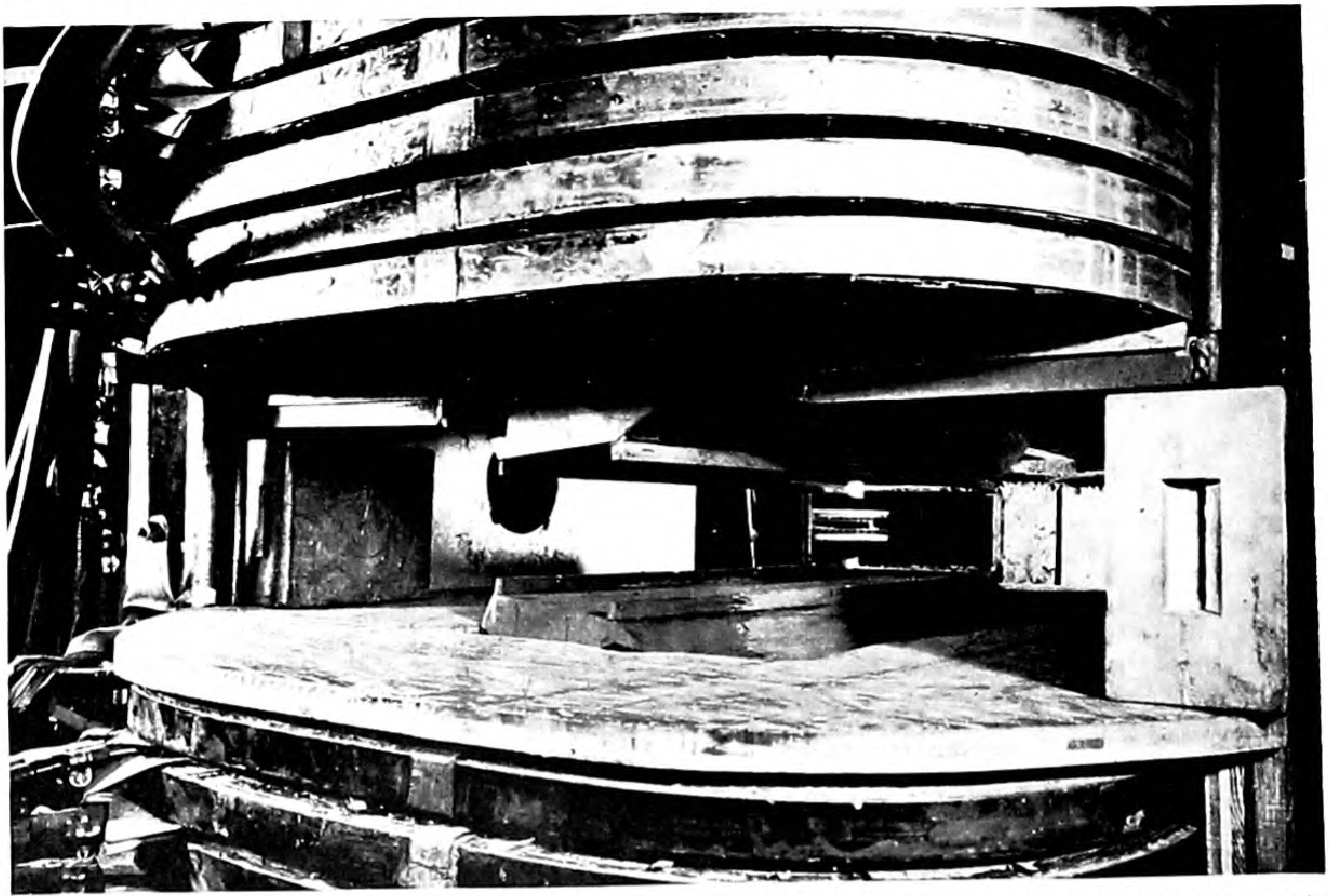
The first such atom to be made was positronium [see "The Ultimate Atom," by H. C. Corben and S. DeBenedetti; SCIENTIFIC AMERICAN, December, 1954]. Positronium is a short-lived, practically weightless atom composed of an electron and a positron—a particle exactly like the electron except that its electric charge is positive instead of negative. The positronium atom is analogous to the simplest ordinary atom, hydrogen, which consists of an electron and a proton; we can consider that in positronium the positron takes the place of the proton.

In this article we shall deal with another type of artificial atom in which we replace an electron with a meson. Here the atom has an ordinary nucleus, consisting of protons and neutrons, but a meson instead of an electron revolves in an orbit around the nucleus. Mesons, as is now well known, are middleweight particles (between the weight of an electron and a proton) which are believed to be connected in some way with the forces inside the atomic nucleus. Since the nature of these forces is the greatest unsolved problem in atomic physics, the mesonic atom is an object of extraordinary interest. What does it tell us about the nucleus?

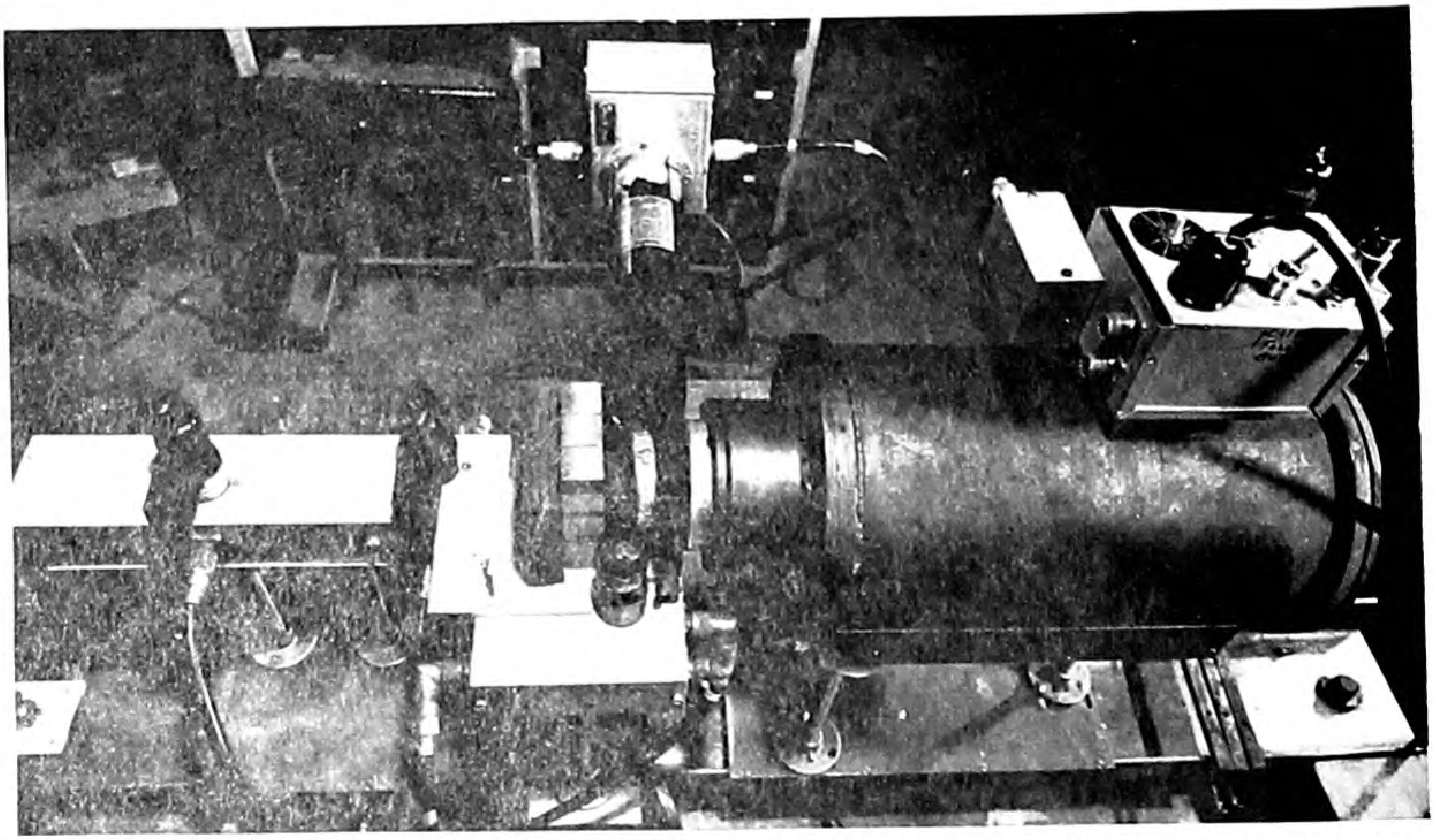
Let us begin by making an inspection tour of the simplest ordinary atom, hydrogen, under the expert guidance of Niels Bohr, who gave us our first idea of its structure. Entering the atom, we find its electron revolving around the heavy positive nucleus, the proton, in a circular orbit about 10^{-8} (a hundred millionth) of a centimeter in diameter. Dr. Bohr points out that if energy is supplied to the atom, the electron may leave this orbit (called the ground state) and jump momentarily to an orbit farther from the nucleus. There is a certain finite number of such orbits available to the electron. It may travel in



ENERGY SPECTRA of electronic and mesonic atoms are compared here. The top three bars locate the emission lines of hydrogen with respect to the visible spectrum (*top*) and to the energy spectrum of hydrogen (*second bar*). The emission lines of electronic neon occur at 100 times the electron voltage of hydrogen. Because the pi meson is 210 times heavier than the electron, the lines of mesonic neon are found at 210 times the electron voltage of electronic neon or 21,000 times the electron voltage of hydrogen.



DEFLECTING MAGNET directs the beam of mesons from the synchrocyclotron to the target. The lucite window of the synchrocyclotron can be seen at right of center through the hole in the 15-foot shielding. The beam travels more than 30 feet in air on its course from the window to the target. The curved metal tracks on the poles of the magnet indicate the deflection of the beam.



MESON TARGET instrumentation is shown here. The meson beam, traveling from left to right in this picture, is detected first by the two counters at left. Next it is slowed by the slabs of copper and beryllium just to left of center. Then it enters the target wedged between two counters at center. X-rays from the target material are measured in the detector, the large tank at right.

any one of these but is never permitted anywhere else within the atom. Soon after its leap to the higher orbit, the electron, being attracted by the positive charge of the nucleus, jumps back to the ground state, in one or more successive steps. Each jump releases energy which we see as light. A familiar example of this light emission is the tube used for luminous signs, where the atoms are excited to higher orbits by an electric discharge. Every atom emits light of characteristic colors. If we analyze with a spectroscope the light of a tube containing hydrogen atoms, we will see a series of sharp lines of different colors, each corresponding to an electronic jump.

We now ask our guide what would happen if the electron of the hydrogen atom were replaced by a negatively charged meson. Dr. Bohr answers that the meson also will be permitted only certain orbits around the nucleus and will emit characteristic radiation at each jump. If the particle is a mu meson, 210 times heavier than the electron, each of its orbits around the nucleus should be 210 times smaller than the corresponding orbit of an electron, and the wavelength of the emitted radiation should be shorter in the same ratio. If the particle is a pi meson, 273 times heavier than the electron, the orbit and radiation wavelength will be reduced by the factor 273.

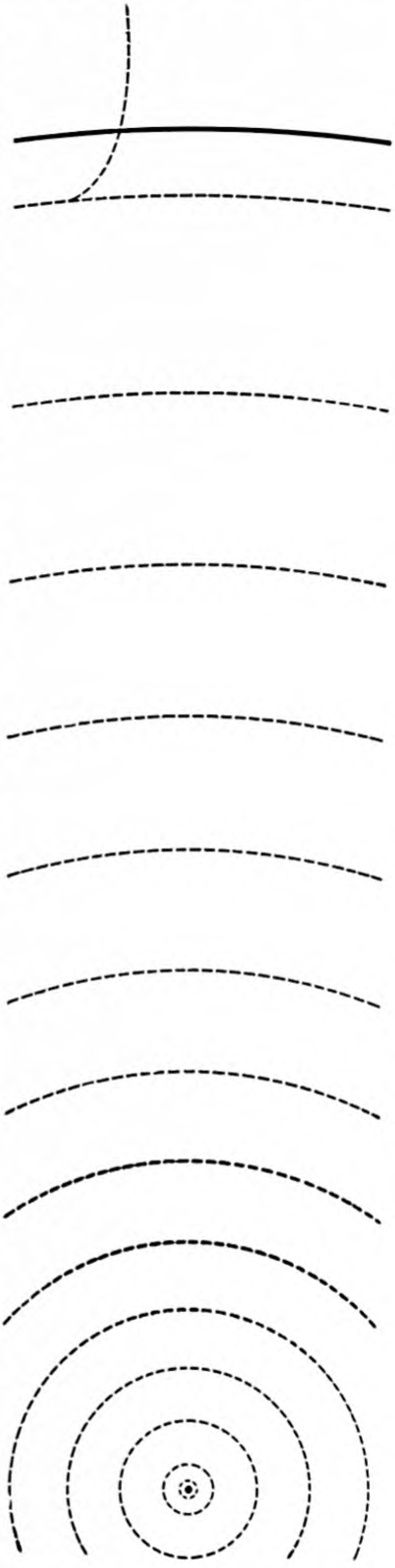
This shortening of the wavelength takes the radiation out of the range of visible light and transfers it to the realm of X-rays. Unfortunately the emission from a mesonic hydrogen atom would be soft (*i.e.*, nonpenetrating) X-rays, which are difficult to study. But a heavier mesonic atom will emit shorter-wave (*i.e.*, more energetic) X-rays. Let us take, for example, the case of neon, an atom containing 10 electrons. Its outermost electronic orbit is about as large as the smallest orbit of hydrogen (10^{-8} of a centimeter in diameter). But its innermost orbit is just 10 times smaller than that of hydrogen. Therefore the smallest orbit of a mu meson replacing an electron in the neon atom would be not 210 but 2,100 times smaller than that of the electron in the unexcited hydrogen atom. There is a corresponding reduction in the wavelength of the radiation emitted: the wavelengths of the radiations from the mesonic jumps of neon should be $210 \times 10 \times 10 = 21{,}000$ times smaller than those of normal hydrogen. We should have no difficulty in detecting X-ray emissions of this energy.

Now that, with the guidance of Dr. Bohr, we have an idea of what to expect of mesonic atoms, let us see how well these predictions are borne out by experiments. The experiments require a synchrocyclotron, to produce beams of negative mesons, and an instrument in which the mesons are captured by atoms and the resulting X-ray emissions are recorded: such an instrument was built by Val Fitch and James Rainwater of Columbia University. The fast mesons emerging from the synchrocyclotron are slowed by passage through a block of solid matter; then, reduced to thermal speed (the ordinary speed of atoms' motions), they enter the material whose atoms are to capture them [*see photograph at bottom on page 50*]. While wandering among these atoms, a meson feels the electrostatic attraction of an atom's positive nucleus and is drawn in to the inner part of the atom, near the nucleus itself. It jumps from one orbit to the next and emits X-rays. The X-rays are registered by a scintillation counter, and their energy, or wavelength, is found by measuring the size of the pulses in the counter.

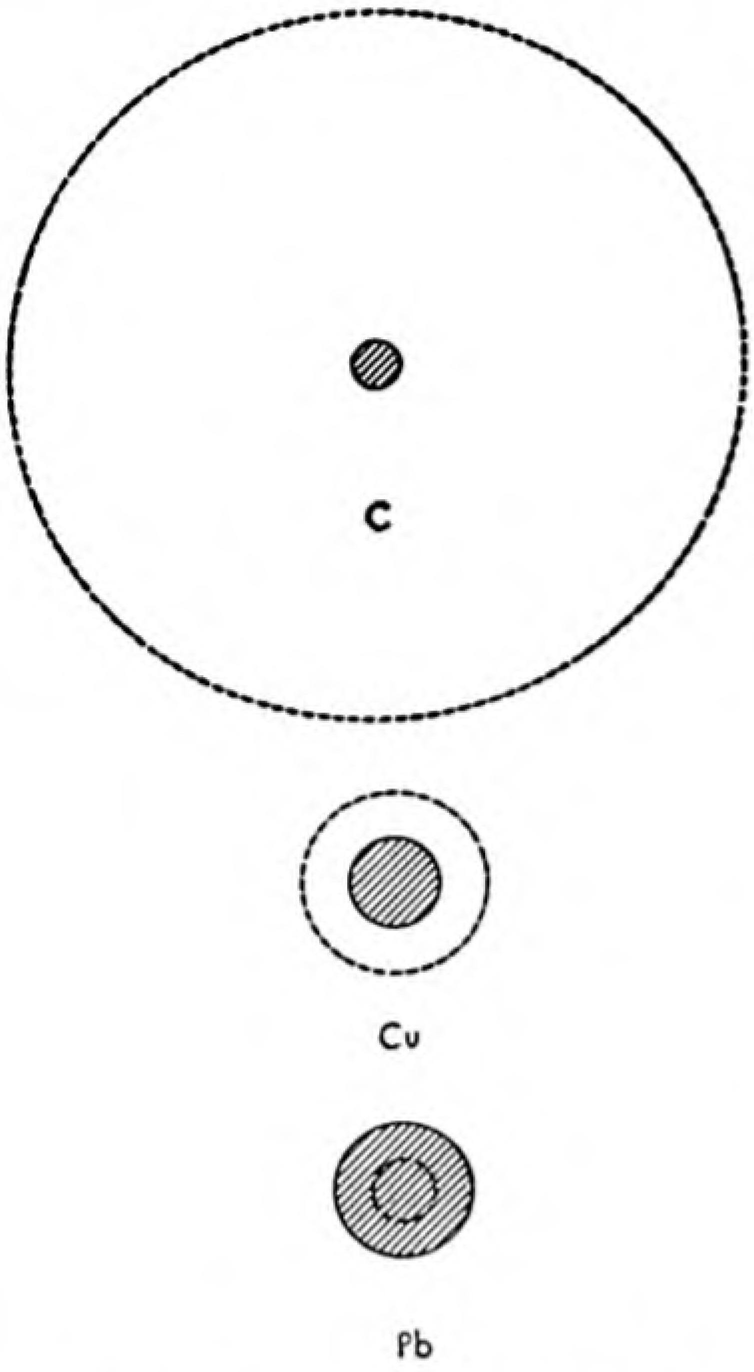
The experiments were first conducted with mu mesons. In the case of relatively light atoms, such as neon or carbon, everything went just as Bohr's theory had predicted. The wavelengths of the X-radiation from the mesonic jumps showed the expected ratio to the wavelengths of light during electronic jumps, as computed from the 210-fold difference in mass between the mu meson and the electron. But when it came to heavy atoms, this regular ratio disappeared. The X-rays emitted by a heavy mesonic atom turned out to be considerably less energetic than expected.

What goes wrong? We begin to get some idea when we look at the dimensions with which we are dealing. Let us take the case of an atom of lead, which has 82 electrons. If we substitute a meson for one of these electrons, then according to Bohr's theory the innermost orbit of the meson should be 82×210 times smaller than the diameter of the hydrogen atom: since the hydrogen diameter is 10^{-8} of a centimeter, the diameter of this orbit is 5.8×10^{-13} of a centimeter. This is a small orbit indeed. Let us look up the diameter of the nucleus of the lead atom. In a table prepared before the experiments of Fitch and Rainwater we find that the diameter of the lead nucleus is given as 17×10^{-13} of a centimeter. In short, the meson's orbit is less than half the size of this nucleus, so that according to our calculation the meson should be revolving within the nucleus!

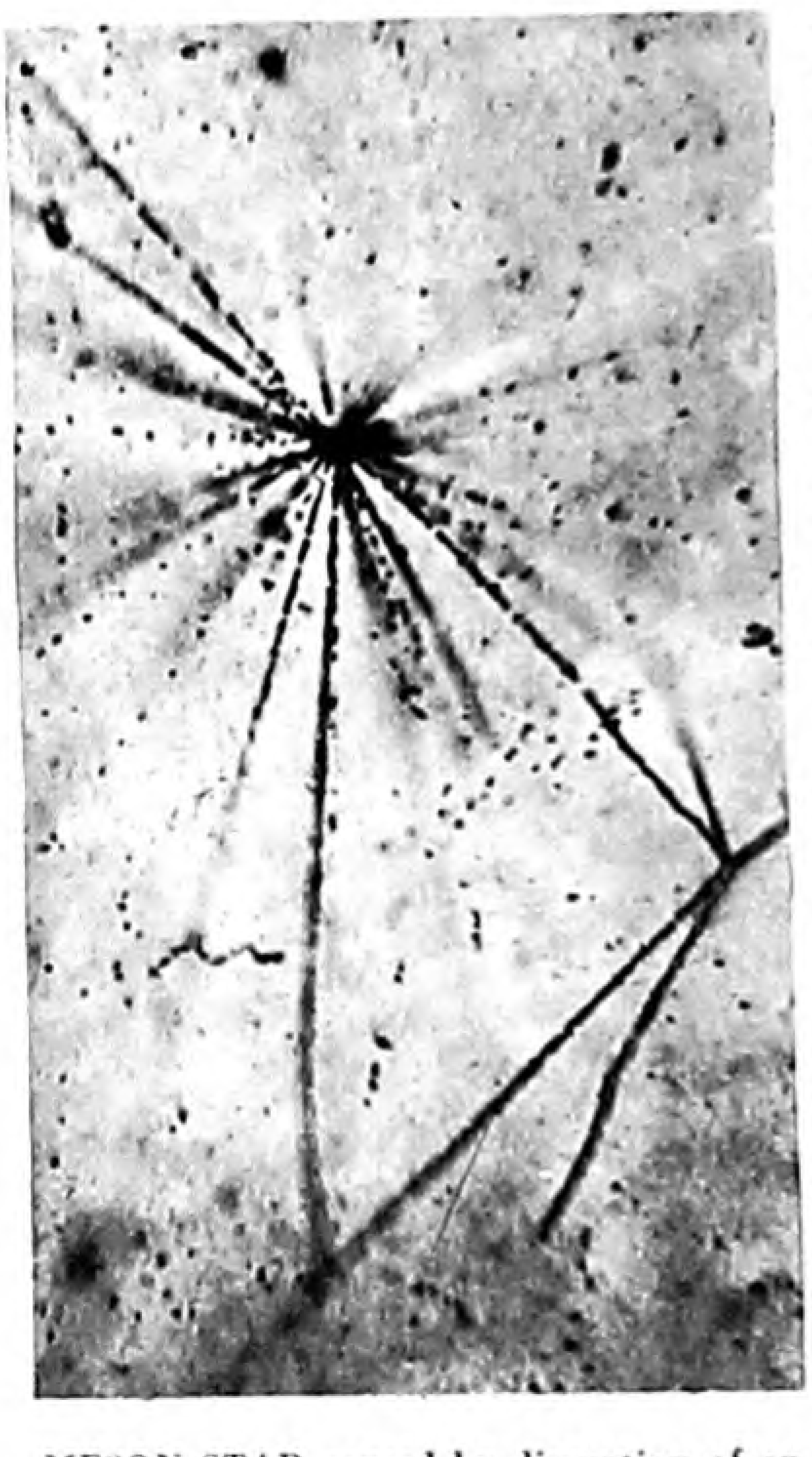
Can this be possible? We have to conclude that, although no such thing is



MESONIC ORBITS of carbon are shown here in rough approximation of their relative distances from the nucleus of the atom at bottom. The heavy arc at top of the diagram locates the innermost electron orbit.



INNERMOST MESONIC ORBITS of carbon, copper and lead are shown here. The nuclei are shown in proportion to size. The innermost orbit of lead is inside the nucleus.



MESON STAR caused by disruption of an atom by a primary cosmic ray appears above center. A pi meson track, traced diagonally to right, terminates in a secondary star.

conceivable in normal atomic physics, it is a possibility in the case of a mesonic atom. While the nucleus of an atom is very dense, density does not necessarily mean opacity, and it is not excluded that the meson may travel freely within the nucleus. The idea of the impenetrability of matter is a macroscopic concept. At the atomic and subatomic level, anything can happen.

And, indeed, it does. The mu meson actually circulates freely within the nucleus (though the pi meson, as we shall see, does not). It makes an enormous number of revolutions—millions of millions—within the nucleus of lead. But all this happens in a hundredth of a millionth of a second, and the meson is then absorbed by the nuclear matter. Upon its absorption the mass of the meson is transformed into energy and the nucleus undergoes a violent explosion.

From the wavelengths of the radiations emitted by mesonic atoms, Fitch and Rainwater were able to calculate the sizes of atomic nuclei. The computations are complex and we need not go into them here, but we can say a few words about the assumptions used. The nucleus is pictured as a cloud of charge—dense but nevertheless perfectly fluid, so that it opposes no resistance to the motion of the meson. For the meson we must use a somewhat more refined model than the old theory of Bohr: it is necessary to take into account the uncertainty principle, according to which a particle cannot be localized as a point moving in a definite orbit but is spread out as if it were a diffuse, jelly-like object of finite extent. When the cloud and the jelly—together with a few more ingredients such as relativity and a spin—are introduced in the Los Alamos electronic computer, out comes the result: the dependence of the wavelength of the X-rays on the size of the nucleus.

About the only thing nuclear physicists thought they knew for sure concerning the nucleus of the atom was its size. But the experiments of Fitch and Rainwater proved that they were wrong even in that. According to their measurements, the nucleus of an atom is only about one half as large in volume as had been thought. This is the first very significant contribution of mesonic atoms to our knowledge of nuclei.

Let us now see what results have been obtained with pi mesons. The pi meson, in contrast to the mu, reacts with nuclear matter very rapidly and much more violently. In a mesonic atom of hydrogen, for instance, the mu meson will revolve peacefully in its orbit around the proton for the comparatively long time (on the atomic scale) of several microseconds; then it decays of its own accord into an electron and neutrinos, just as if the proton were not there at all. On the other hand, a pi meson in such an atom barely reaches the lowest orbit before it is gobbled up by the proton. Its lifetime in the atom is a million times shorter than that of a mu meson. When the negative pi meson reacts with the positive proton, they neutralize each other's electric charge and become neutral particles.

In a heavier atom the phenomenon is even more spectacular. In neon the pi meson does not reach the lowest orbit: it is eaten up by the nucleus when it arrives at the next-to-lowest. The greediness of nuclei for the pi meson is almost incredible. In a heavy atom such as lead —in which the mu meson can travel almost undisturbed within the nucleus— the pi is captured when still in the fifth or sixth orbit away from the nucleus— orbits whose diameter is at least 10 times larger than the nucleus itself. The experimental evidence for this behavior is the absence of the last X-ray lines.

After a pi meson is captured by the nucleus, it disappears entirely. As in the case of the mu meson, its mass is transformed into energy: the nucleus explodes and breaks into many pieces. In a photographic emulsion the pieces flying away leave a developable image in the characteristic form of a star.

The fact that the nucleus captures the meson from a faraway orbit does not necessarily mean that they are brought together by a force of attraction. As we have mentioned before, the whole picture of orbits is a convenient oversimplification: the particle which in Bohr's theory is supposed to travel in an orbit is really smeared out over most of the atom, and in a sense it touches the nucleus. Thus no special forces are needed to get the meson to the nucleus.

What light can mesonic atoms throw on our main problem—the forces that hold the protons and neutrons together in the nucleus of an atom? Most nuclear physicists now believe that the key to this puzzle lies in the pi meson (which goes under the name of "nuclear glue" even in the daily press). It seems evident that there are strong forces, other than electrostatic, between pi mesons and atomic nuclei. For one thing, a beam of pi mesons behaves differently from one of mu mesons when it bombards matter. The pi mesons will change their

direction and scatter much more than the mu: since the electric forces are the same in the two cases, the difference is attributed to specific nuclear forces acting on the pi meson.

The scattering of pi mesons was first studied with some care by the late Enrico Fermi and his collaborators at the University of Chicago. One of the points they could not determine was the sign of the force: whether it was a force of attraction or repulsion. Now, with the mesonic atom, it has been possible to test this question. If the force is attractive, we should expect the pi meson to be brought closer to the nucleus than if it were repulsive, and the issue should be decidable by examination of the wavelengths of X-ray emission. With this in mind, a careful measurement of the X-rays from pi-mesonic atoms was performed at the Carnegie Institute of Technology by Martin and Mary Stearns Larry Leipuner and the author. The results showed that the nuclear force on the pi meson is repulsive. This conclusion, which has since been verified by certain detailed features of the scattering, need not upset our ideas about the nucleus itself. The forces between the neutron and the proton can still be attractive, and there is no danger that atomic nuclei will come apart!

In spite of its obvious involvement, no quantitative relation between the behavior of the pi meson and the nuclear forces proper has yet been found. Perhaps other mesons and new particles which are continually turning up have something to do with these forces. At any rate, the mesonic atom offers a new approach which is full of promise. As soon as sufficiently intense beams of the newer mesons (tau, k, etc.) become available from the bigger accelerators, we may hope to build new atoms with them and perhaps learn more about the properties of the nucleus.

The Author

SERGIO DEBENEDETTI is professor of physics at the Carnegie Institute of Technology. He was born in Florence in 1912, and studied in the Laboratory of Arcetri, near the hill where Galileo died. He took his Ph.D. at Florence, where he was associated with Bruno Rossi, the cosmic ray expert. In 1938 he left Italy for the Curie Laboratory in Paris, and there developed an interest in positrons. DeBenedetti has been in the U.S. since 1940; in 1946–1948 he was principal physicist at the Clinton Laboratories in Oak Ridge. His interest in elementary particles has led him to the special, and enjoyable, occupation of making atoms with new particles. As for hobbies, he says: "The thing I really like is to see the world, and with the excuse of physics and of cosmic rays I have touched all continents but Asia and Australia, a situation which I intend to remedy before too long. Of course I am interested in world affairs. I also like to draw and paint."

Bibliography

ATOMIC AND NUCLEAR PHYSICS. Robert S. Shankland. The Macmillan Co., 1955.

MESONS, VOLUME II. H. Bethe and F. deHoffmann. Row, Peterson & Company, 1955.

SCIENTIFIC AMERICAN April 1953, Vol. 188, No. 4, pp. 57-64 OFFPRINT 208

FIELD THEORY

by Freeman J. Dyson

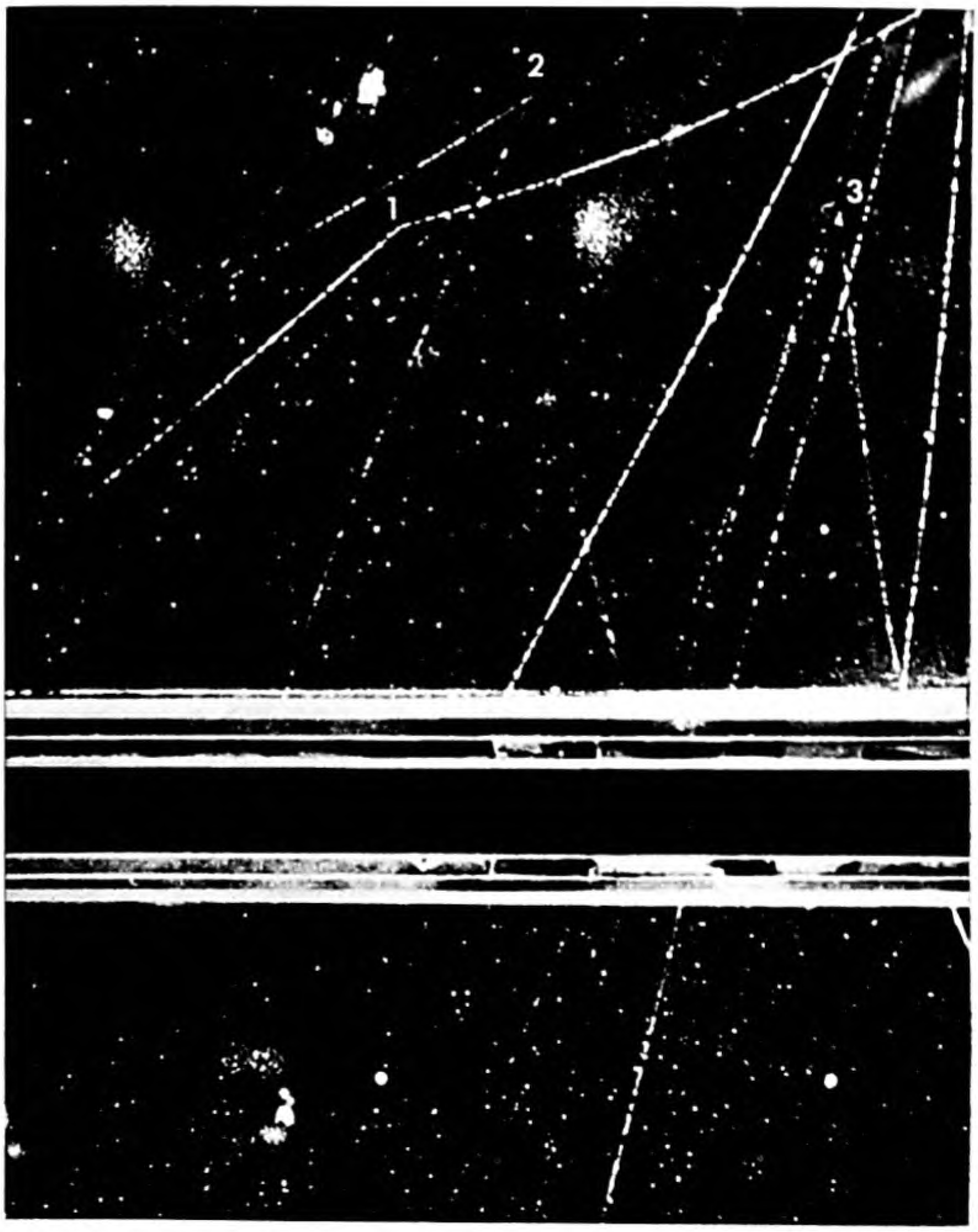
The physics of the 19th century discovered "classical" fields; modern physics deals with quantum fields. What is the nature of these quantum fields, and what place do they occupy in our present view of reality?

"IT IS perhaps surprising that no new meson was reported during the symposium, though almost a month had passed since a previous meeting of nuclear physicists in Copenhagen."

This learned joke in the British journal *Nature*, commenting on an international physics conference last summer, sums up very well the present chaotic situation in theoretical physics. We have become accustomed during the last few years to the discovery of new particles. About 20 different kinds are now known. Everybody expects that many more will be discovered as experimental techniques are improved. Yet nobody has had any success in classifying the known particles, or in predicting the properties of unknown ones. Nobody understands why such and such particles exist, why they have the particular masses that are observed or why some of them strongly interact and some do not.

How do the theoretical physicists spend their time, if they are not able to attack the fundamental problem of the nature of elementary particles? Of what use can the existing atomic theories be, if they do not throw light on this basic problem? These awkward questions are asked rather frequently when experimental and theoretical physicists come together. I shall try to answer them and to explain why we theoretical physicists believe that our theories are useful even though there is so much we do not understand.

First it is necessary to make one point clear: there *is* an official and generally accepted theory of elementary particles, known as the "quantum field theory." While theoretical physicists often disagree about the finer details of the theory, and especially about the way in which it should be applied to practical problems, the great majority of them agree that the theory in its main features is correct. The minority who reject the theory, although led by the great names of Albert Einstein and P. A. M. Dirac, do not yet have any workable alternative to put in its place. In this article I shall adopt the point of view of the majority. When I talk about the concept of field, I mean specifically the concept as it is used in the present-day official quantum field theory. The majority believes that this concept is so useful and illuminating that it will survive the changes and revolutions which the theory will inevitably undergo in the future. Henceforward I shall omit the phrase "in the



NEUTRAL V-PARTICLES gave rise to the three V-shaped tracks in this cloud-chamber photograph by R. B. Leighton of the California Institute of Technology. Below the center of the photograph is the edge of a lead plate.

opinion of the majority," or "in my opinion," which strictly ought to stand at the beginning of every sentence.

A Descriptive Theory

It is important to make a second general remark about the theory at the outset. This concerns the failure of the theory to give us an understanding of why the observed elementary particles exist and no others. The point is that the theory is in its nature descriptive and not explanatory. It describes how elementary particles behave; it does not attempt to explain why they behave so. To draw an analogy from a familiar branch of science, the function of chemistry as it existed before 1900 was to describe precisely the properties of the chemical elements and their interactions. Chemistry described how the elements behave; it did not try to explain why a particular set of elements, each with its particular properties, exists. To answer the question "why," completely new sciences were needed: atomic and nuclear physics. Looking backward, it is now clear that 19th-century chemists were right to concentrate on the "how" and to ignore the "why." They did not have the tools to begin to discuss intelligently the reasons for the individualities of the elements. They had to spend a hundred years building up a good quantitative descriptive theory before they could go further. And the result of their labors—the classical science of chemistry—was not destroyed or superseded by the later insight that atomic physics gave.

The quantum field theory treats elementary particles just as 19th-century chemists treated the elements. The theory starts from the existence of a specified list of elementary particles, with specified masses, spins, charges and specified interactions with one another. All these data are put into the theory at the beginning. The purpose of the theory is simply to deduce from this information what will happen if particle A is fired at particle B with a given velocity. We are not yet sure whether the theory will be able to fulfill even this modest purpose completely. Many technical difficulties have still to be overcome. One of the difficulties is that we do not yet have the complete list of elementary particles. Nevertheless the successes of the theory in describing experimental results have been striking. It seems likely that the theory in something like its present form will describe accurately a very wide range of possible experiments. This is the most that we would wish to claim for it.

Our justification for concentrating attention so heavily on the existing theory, with its many arbitrary assumptions, is the belief that a working descriptive theory of elementary particles must be established before we can expect to reach a more complete understanding at a deeper level. The numerous attempts to by-pass the historical process, and to understand the elementary particles on the basis of general principles without waiting for a descriptive theory, have been as unsuccessful as they were ambitious. In fact, the more ambitious they are, the more unsuccessful. These attempts seem to be on a level with the famous 19th-century attempts to explain atoms as "vortices in the ether."

Classical Fields

Physicists talk about two kinds of fields: classical fields and quantum fields. Actually we believe that all fields in nature are quantum fields. A classical field is just a special large-scale manifestation of a quantum field. But since classical fields were discovered first and are easier to understand, it is necessary to say what we mean by a classical field first, and go on to talk about quantum fields later.

A classical field is a kind of tension or stress which can exist in empty space in the absence of matter. It reveals itself by producing forces, which act on any material objects that happen to lie in the space the field occupies. The standard examples of classical fields are the electric and magnetic fields, which push and pull electrically charged objects and magnetized objects respectively. Michael Faraday discovered that these two fields also exert effects on each other. He found that a changing magnetic field produces electric forces (an effect now known as induction), and his finding made possible the development of practical electric generators. Later the exact laws of behavior of electric and magnetic fields were formulated mathematically by James Clerk Maxwell. He found that in any space where a changing magnetic field exists, an electric field must exist also, and *vice versa*. In order to describe completely the state of the fields in a given region of space, it is necessary to specify the strength and the direction of both the electric and magnetic fields at every point of the region separately. This is the characteristic mathematical property of a classical field: it is an undefined something which exists throughout a volume of space and which is described by sets of numbers, each set denoting the field strength and direction at a single point in the space.

Maxwell was the first to realize that electric and magnetic fields could exist not only near charges and magnets but also in free space completely disconnected from material objects. From his equations he deduced that in empty space such fields would travel with the velocity of light. Hence he made the epoch-making guess that light consists of traveling electromagnetic fields. We now know that his guess was correct, and we are even able to manufacture traveling electromagnetic fields ourselves and use them for various purposes. These artificial traveling fields we call radio.

Another example of a classical field is the gravitational field. This has the special property that it acts on all material objects in a given region of space. It is very difficult to experiment with, because the gravitational field produced by any object of convenient laboratory size is absurdly weak. For this reason we have never been able to detect any effects of freely traveling gravitational waves, which presumably exist in the neighborhood of a rapidly oscillating mass. It is also impossible to measure any possible interactions of the gravita-

NAME	SYMBOL
PHOTON	γ
GRAVITON	G
NEUTRINO	ν
ELECTRON	e
POSITRON	p
POSITIVE MU MESON	μ^+
NEGATIVE MU MESON	μ^-
NEUTRAL PI MESON	π^0
POSITIVE PI MESON	π^+
NEGATIVE PI MESON	π^-
ZETA MESON?	ζ
NEUTRAL V-PARTICLE	V_2^0
TAU MESON	τ
KAPPA MESON	κ
POSITIVE CHI MESON	χ^+
NEGATIVE CHI MESON	χ^-
PROTON	P
NEUTRON	N
NEUTRAL V-PARTICLE	V_1^0
POSITIVE V-PARTICLE?	V^+

CHART of the fundamental particles has been revised since a similar chart appeared in this magazine for

CHARGE	MASS	SPIN	STATISTICS	LIFETIME (SECONDS)	DECAY SCHEME
0	0	1	BOSE-EINSTEIN	STABLE	
0	0	2	BOSE-EINSTEIN	STABLE	
0	0	½	FERMI-DIRAC	STABLE	
−	1	½	FERMI-DIRAC	STABLE	
+	1	½	FERMI-DIRAC	STABLE	
+	210	½	FERMI-DIRAC	2.1×10^{-6}	$\mu^+ \rightarrow p + 2\nu$
−	210	½	FERMI-DIRAC	2.1×10^{-6}	$\mu^- \rightarrow e + 2\nu$
0	265	0	BOSE-EINSTEIN	10^{-15}	$\pi^0 \rightarrow 2\gamma$
+	276	0	BOSE-EINSTEIN	2.6×10^{-8}	$\pi^+ \rightarrow \mu^+ + \nu$
−	276	0	BOSE-EINSTEIN	2.6×10^{-8}	$\pi^- \rightarrow \mu^- + \nu$
±	550	?	?	10^{-12}	$\zeta \rightarrow \pi + ?$
0	850	?	?	10^{-10}	$V_2^0 \rightarrow \pi^+ + \pi^- + ?$
±	975	?	BOSE-EINSTEIN	10^{-8}	$\tau \rightarrow 3\pi$
±	1100	?	?	?	$\kappa \rightarrow \mu + ?$
+	1400	?	?	10^{-9}	$\chi^+ \rightarrow \pi^+ + ?$
−	1400	?	?	10^{-9}	$\chi^- \rightarrow \pi^- + ?$
+	1836	½	FERMI-DIRAC	STABLE	
0	1838.5	½	FERMI-DIRAC	750	$N \rightarrow P + e + \nu$
0	2190	?	FERMI-DIRAC	3×10^{-10}	$V_1^0 \rightarrow P + \pi^-$
+	2200	?	?	10^{-9}	$V^+ \rightarrow P + ?$

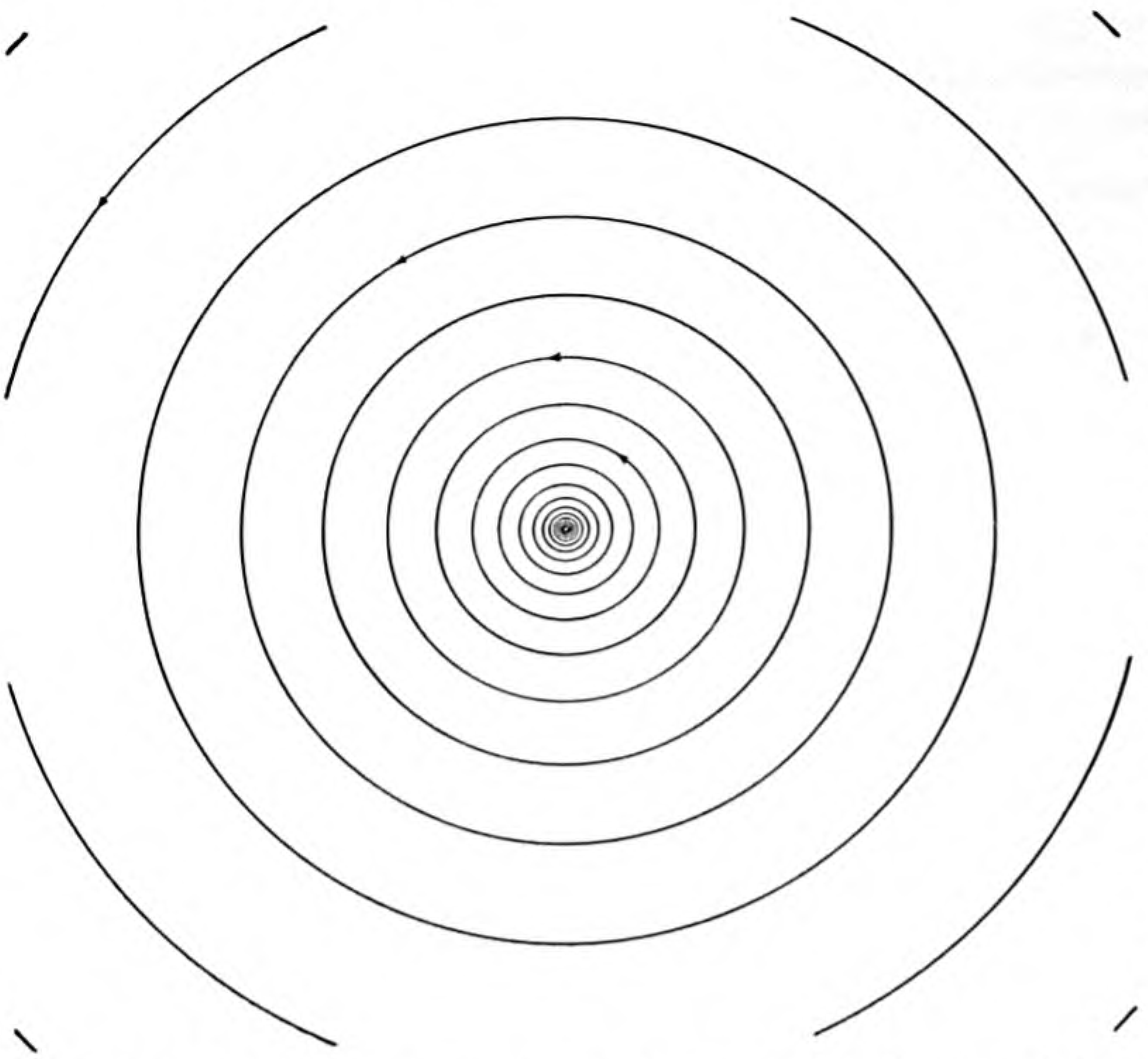
January, 1952. Among the changes are the addition of the chi mesons and a second variety of neutral V-particle. The particles are listed in the order of their mass. The existence of the zeta meson and the positive V-particle is doubtful. The masses and the lifetimes of all the newer particles listed in the chart are approximate.

tional and electromagnetic fields. This is most unfortunate, and it is the main reason why we know so much less about gravitation than about the other fields.

A Model of a Field

What, then, is the picture we have in mind when we try to visualize a classical field? Characteristically, modern physicists do not seriously try to visualize the objects they discuss. In the 19th century it was different. Then it seemed that the universe was built of solid mechanical objects, and that to understand an electric field it was necessary to visualize the field as a mechanical stress in a material substance. It was possible, indeed, to visualize electric and magnetic fields in this way. To do so men imagined a material substance called the ether, which was supposed to fill the whole of space and carry the electric and magnetic stresses. But as the theory was developed, the properties of the ether became more and more extraordinary and self-contradictory. Einstein in 1905 finally abandoned the ether and proposed a new and simple version of the Maxwell theory in which the ether was never once mentioned. Since 1905 the idea that everything in the universe should be visualized mechanically has gradually become ridiculous. We now find that mechanical objects themselves are composed of atoms held together by electric fields, and therefore it makes no sense to try to explain the electric fields in terms of mechanical models.

It is still convenient sometimes to make a mental picture of an electric field. For example, we may think of it as a flowing liquid which fills a given space and which at each point has a certain velocity (strength) and direction of flow. But nobody nowadays imagines that the liquid really exists or that it explains the behavior of the field. The flowing liquid is just a model—a convenient way to express our knowledge about the field in concrete terms. It is a good model only so long as we remember not to take it too seriously. We must not, for example, expect that the equations of motions of the electric field will be the same as those of any self-respecting liquid. To a modern physicist the electric field is a fundamental concept which cannot be reduced to anything simpler. It is a unique something with a



CLASSICAL FIELD is schematically depicted as a series of concentric rings around a point. These rings might show the magnetic field around an electric current traveling along a wire perpendicular to the surface of this page.

set of known properties, and that is all there is to it. This being understood, the reader may safely think of the flowing liquid as a fairly accurate representation of what we mean by a classical electric field. The electric and magnetic fields must then be pictured as two different liquids, both filling the whole of space, moving separately and interpenetrating each other freely. At each point there are two velocities, representing the strengths of the electric and magnetic components of the total electromagnetic field.

It is characteristic of a classical field that its strength at a given point varies smoothly as the point moves around in space. Therefore the liquid model must be imagined as an ideal liquid, not composed of atoms but filling all space uniformly and having a well-defined velocity at every point.

The new idea that Einstein introduced in 1905, and that killed the ether, was the principle of relativity. This principle states that the properties of empty space are always the same, regardless of the velocity with which an experimenter is moving through it. Thus even if there is a material ether filling space, the experimenter is unable to measure the velocity of himself relative to it; for all practical purposes the ether is unobservable. All that we can certainly say about it is that if it does exist, it is of no interest to us. Our picture of the world becomes much simpler if we abandon the ether and speak only about electric and magnetic fields in empty space.

Einstein made a complete theory of the classical electromagnetic field and its interactions with matter, using the principle of relativity as his starting point. In 1916 he extended the idea of relativity to construct his theory of the classical gravitational field. These theories stand today substantially as Einstein left them.

The classical field theories of Einstein —electromagnetic and gravitational—together give us a satisfactory explanation of all large-scale physical phenomena. That is to say, they explain everything in the physical world that can be explained without bringing into view the fact that the world is built of elementary particles. There is every reason to believe that the classical field theories are correct so long as we are talking about objects much bigger and heavier than a single atom. But they fail completely to describe the behavior of individual atoms and particles. To understand the small-scale side of physics, physicists had to invent quantum mechanics and the idea of a quantum field.

Quantum Fields

Unfortunately the quantum field is even more difficult to visualize than the classical field. The basic axiom of quantum mechanics is the uncertainty principle. This says that the more closely we look at any object, the more the object is disturbed by our looking at it, and the less we can know about the subsequent state of the object. Another less precise way of expressing the same principle is this: All objects of atomic size fluctuate continually; they cannot maintain a precisely defined position for a finite length of time. Their quantum fluctuations are never precisely predictable, and the laws of quantum mechanics tell us only the statistical behavior of the fluctuations when averaged over a long time. The universal existence of these fluctuations, and the general correctness of the laws of quantum mechanics, have been verified by a wealth of experiments during the last 30 years.

How do the quantum fluctuations affect the classical field? The answer is: not at all. The fluctuations are not observable with any ordinary large-scale equipment, for they average out to produce no effect on these instruments. Looked at with large-scale apparatus, the quantum field behaves exactly like a classical field. Only when we measure the effects of an electromagnetic field on a single atom do the quantum fluctuations of the field become noticeable.

The physicists Willis Lamb and Robert C. Retherford at Columbia University have observed the effects of electromagnetic fields on single hydrogen atoms with a piece of apparatus known to radar experts as a microwave cavity resonator [see "Radio Waves and Matter," by Harry M. Davis; SCIENTIFIC AMERICAN, September, 1948]. Using the techniques of microwave spectroscopy, they were able to measure the effects of the fields with great accuracy. The effect of the quantum fluctuations, itself a small part of the total effect of the fields, was measured to an accuracy better than one part in a thousand, and within this margin of possible error the effect agreed with the conclusions of the quantum field theory. The Lamb-Retherford experiment is the strongest evidence we have for believing that our picture of the quantum field is correct in detail.

At the risk of making some professional quantum theoreticians turn pale, I shall describe a mechanical model which may give some idea of the nature of a quantum field. Imagine the flowing liquid which served as a model for a classical electric field. But suppose that the flow, instead of being smooth, is turbulent, like the wake of an ocean liner. Superimposed on the steady average motion there is a tremendous confusion of eddies, of all different sizes and overlapping and mingling with one another. In any small region of the liquid the velocity continually fluctuates, in a more or less random way. The smaller the region, the wilder and more rapid are the velocity fluctuations. In a real liquid these fluctuations are finally limited by two factors: (1) the viscosity,

or stickiness, of the liquid, which damps out the turbulent motions, and (2) the atomic structure of the liquid, which sets a minimum size for the eddies, since it is meaningless to talk about eddies containing only a few atoms. In our model of the quantum field, however, we assume that neither of these factors operates. There is no dissipation of energy by viscosity, or any minimum size of eddies. Consequently the velocity in a given region can continue to fluctuate without diminution forever, and the fluctuations grow more and more intense without limit as the size of the region is reduced.

The model does not describe correctly the detailed quantum-mechanical properties of a quantum field; no classical model can do that. But it does seem to me to give a reasonably valid picture of the general appearance of the thing. In particular, the model makes clear that it is strictly meaningless to speak about the velocity of the liquid at any specific point. The fluctuations in the neighborhood of the point become infinitely large as the neighborhood becomes smaller, and so the velocity at the point itself has no meaning. The only quantities that have meaning are averaged velocities, taken over a given region of space and over a given time interval. This property of the model is a true representation of a property of a quantum field. The strength of a quantum field at a point can never be measured. The whole quantum field theory is a theory of the behavior of field strengths averaged over finite regions of space and time.

The Particles Emerge

Now comes the climax of the story. We have put into the theory of the quantum field two big ideas: the idea of quantum mechanics and the idea of relativity. These two ideas force us to construct a mathematical theory which in its main lines is fixed; the only freedom left to us is in matters of detail. When we deduce the consequences of this mathematical theory, we find that a miracle has occurred: automatically there emerges a third big idea—that the world is built of elementary particles.

This idea is a consequence of the fact that in a quantum field energy can exist only in discrete units, which we call quanta. When we work out the theory of these quanta in detail, we find that they have precisely the properties of the elementary particles that we observe in the world around us.

It is not possible, in an article such as this, to explain how the elementary particles arise mathematically out of the fluctuations of a field. It cannot be understood by thinking about turbulent liquids or any classical model. All I can say here is that it happens. And it is the basic permanent reason for believing that the concept of a quantum field is a valid concept and will survive any changes that may later be made in matters of detail.

The picture of the world that we have finally reached is the following: Some 10 or 20 qualitatively different quantum fields exist. Each fills the whole of space and has its own particular properties. There is nothing else except these fields; the whole of the material universe is built of them. Between various pairs of the fields there are various kinds of interaction. Each field manifests itself as a type of elementary particle. The particles of a given type are always completely identical and indistinguishable. The number of particles of a given type is not fixed, for particles are constantly being created or annihilated or transmuted into one another. The properties of the interactions determine the rules for creation and transmutation of particles.

In this picture of the world the electromagnetic field appears on an exactly equal footing with the other fields. The particle corresponding to it is the light quantum, or photon. The photon appears to be different from other elementary particles only because its laws of interaction make it especially easy to create and annihilate. So the photon appears to be less permanent than, for example, the electron. But this is only a difference of degree; all particles, including the electron, can be rapidly annihilated under suitable conditions.

The elementary particle corresponding to the gravitational field has been named the graviton. There can be little doubt that in a formal mathematical sense the graviton exists. However, nobody has ever observed an individual graviton. Because of the extreme weakness of the gravitational interaction, in practice only large masses produce observable gravitational effects. In the case of large masses, the number of gravitons involved in the interaction is very large, and the field behaves like a classical field. Consequently, many physicists believe that the individual graviton never will be observed. Whether the graviton has a real existence is one of the most important open questions in physics.

The electromagnetic and gravitational fields have one essential property in common. They are long-range fields which make their effects felt over great distances. This is connected with the fact that the photon and the graviton are particles which have no rest-mass and always travel at a fixed velocity—the velocity of light. Almost all other fields in nature have a short range, less than the size of an atom, and their effects cannot be felt beyond this distance. The short-range fields cannot be detected in a classical way by measuring their effects on large objects. They never behave like classical fields in any experimental situation. This is why, for example, the field corresponding to the electron was never recognized as a field until the quantum field theory was developed. And even now the electron field seems more peculiar and foreign to us than the electromagnetic field. Fundamentally the two are very similar. The main difference between them is the short range of the electron field, which has the consequence that the electron possesses a rest-mass and can travel with any velocity not exceeding the velocity of light. Most of the other known particles—protons, neutrons, the many varieties of mesons—also have a rest-mass and are associated with short-range fields.

Positive and Negative

Perhaps the most spectacular success of the quantum field theory is in its treatment of charged fields. According to the theory, a quantum field may or may not carry an electric charge. For example, the electron field carries a charge, while the electromagnetic field does not. The theory automatically predicts that any charged field must be represented by two types of particle, precisely alike in all respects except that one has a positive charge and the other negative. The theory also predicts that under suitable conditions a pair of such particles, one positively and one negatively charged, can be created or annihilated together in a single event. All these predictions of the theory have been completely confirmed in the case of the electron field. There exists a particle, the positron, which is exactly like an electron except that it has the opposite charge. It has also been proved that there are at least two varieties of meson that exist in positive and negative forms. The theory predicts that there should be an antiproton: a particle negatively charged but otherwise identical with a proton. The antiproton has not yet been detected. It presents an outstanding challenge to experimental physicists to discover it, or to theoretical physicists to explain why it should not exist.

Even to a hardened theoretical physicist it remains perpetually astonishing that our solid world of trees and stones can be built of quantum fields and nothing else. The quantum field seems far too fluid and insubstantial to be the basic stuff of the universe. Yet we have learned gradually to accept the fact that the laws of quantum mechanics impose their own peculiar rigidity upon the fields they govern, a rigidity which is alien to our intuitive conceptions but which nonetheless effectively holds the earth in place. We have learned to apply, both to ourselves and to our subject, the words of Robert Bridges:

Our stability is but balance, and
our wisdom lies
In masterful administration of the
unforeseen.

The Author

FREEMAN J. DYSON is one of the chief architects of the theory of the quantum electromagnetic field. He concisely outlines his own career thus: "Born in England in 1923. Son of Sir George Dyson, professional musician. Started off with an appetite for mathematics and astronomy from the age of six. Studied mainly mathematics at Cambridge University. Spent the last two years of the war at headquarters of R.A.F. Bomber Command doing operations research. Investigated causes of bomber losses in night operations. Found this a frustrating experience, scientific honesty only rarely being allowed to prevail over political expediency. Decided to make a fresh start after reading the Smyth Report in the fall of 1945, thinking that physics would be the major stream of scientific progress during the next 25 years. Also encouraged to become a physicist by the discovery that physics was in more of a mess than mathematics or astronomy. Came to America with a Commonwealth Fund Fellowship in 1947 and learnt most of the physics I know from Professors Bethe and Feynman at Cornell University. Was lucky to arrive and start research in the exciting days of 1947 when the Lamb-Retherford experiment was new and Bethe and Feynman were busy understanding it. This determined the direction of all my subsequent work. Studied for a time at the Institute for Advanced Study, Princeton, and married one of the mathematicians there. I am now back at Cornell as a professor of physics."

Bibliography

QUANTUM ELECTRODYNAMICS. F. J. Dyson in *Physics Today*. September, 1952.

THE STRUCTURE OF MATTER. Francis Owen Rice and Edward Teller. John Wiley & Sons, Inc., 1949.

SPECIAL NOTE TO TEACHERS: Each article in this volume, plus more than 660 others, is available as a separate, self-bound SCIENTIFIC AMERICAN Offprint. Offprints may be ordered in any combination and in any quantity. Teachers who want to adopt articles for their courses, therefore, can ensure that each student has his own set. Students' sets are collated by the publisher before shipment.

SCIENTIFIC AMERICAN April 1950, Vol. 188, No. 4, pp. 13-17 OFFPRINT 209

ON THE GENERALIZED THEORY OF GRAVITATION

by Albert Einstein

An account of the extension of the general theory of relativity against its historical and philosophical background.

THE editors of SCIENTIFIC AMERICAN have asked me to write about my recent work which has just been published. It is a mathematical investigation concerning the foundations of field physics.

Some readers may be puzzled: Didn't we learn all about the foundations of physics when we were still at school? The answer is "yes" or "no," depending on the interpretation. We have become acquainted with concepts and general relations that enable us to comprehend an immense range of experiences and make them accessible to mathematical treatment. In a certain sense these concepts and relations are probably even final. This is true, for example, of the laws of light refraction, of the relations of classical thermodynamics as far as it is based on the concepts of pressure, volume, temperature, heat and work, and of the hypothesis of the non-existence of a perpetual motion machine.

What, then, impels us to devise theory after theory? Why do we devise theories at all? The answer to the latter question is simply: Because we enjoy "comprehending," *i.e.*, reducing phenomena by the process of logic to something already known or (apparently) evident. New theories are first of all necessary when we encounter new facts which cannot be "explained" by existing theories. But this motivation for setting up new theories is, so to speak, trivial, imposed from without. There is another, more subtle motive of no less importance. This is the striving toward unification and simplification of the premises of the theory as a whole (*i.e.*, Mach's principle of economy, interpreted as a logical principle).

There exists a passion for comprehension, just as there exists a passion for music. That passion is rather common in children, but gets lost in most people later on. Without this passion, there would be neither mathematics nor natural science. Time and again the passion for understanding has led to the illusion that man is able to comprehend the objective world rationally, by pure thought, without any empirical foundations—in short, by metaphysics. I believe that every true theorist is a kind of tamed metaphysicist, no matter how pure a "positivist" he may fancy himself. The metaphysicist believes that the logically simple is also the real. The tamed metaphysicist believes that not all that is logically simple is embodied in experienced reality, but that the totality of all sensory experience can be "comprehended" on the basis of a conceptual system built on premises of great simplicity. The skeptic will say that this is a "miracle creed." Admittedly so, but it is a miracle creed which has been borne out to an amazing extent by the development of science.

The rise of atomism is a good example. How may Leucippus have conceived this bold idea? When water freezes and becomes ice—apparently something entirely different from water—why is it that the thawing of the ice forms something which seems indistinguishable from the original water? Leucippus is puzzled and looks for an "explanation." He is driven to the conclusion that in these transitions the "essence" of the thing has not changed at all. Maybe the thing consists of immutable particles and the change is only a change in their spatial arrangement. Could it not be that the same is true of all material objects which emerge again and again with nearly identical qualities?

This idea is not entirely lost during the long hibernation of occidental thought. Two thousand years after Leucippus, Bernoulli wonders why gas exerts pressure on the walls of a container. Should this be "explained" by mutual repulsion of the parts of the gas, in the sense of Newtonian mechanics? This hypothesis appears absurd, for the gas pressure depends on the temperature, all other things being equal. To assume that the Newtonian forces of interaction depend on temperature is contrary to the spirit of Newtonian mechanics. Since Bernoulli is aware of the concept of atomism, he is bound to conclude that the atoms (or molecules) collide with the walls of the container and in doing so exert pressure. After all, one has to assume that atoms are in motion; how else can one account for the varying temperature of gases?

A simple mechanical consideration shows that this pressure depends only on the kinetic energy of the particles and on their density in space. This should have led the physicists of that age to the conclusion that heat consists in random motion of the atoms. Had they taken this consideration as seriously as it deserved to be taken, the development of the theory of heat—in particular the discovery of the equivalence of heat and mechanical

energy—would have been considerably facilitated.

This example is meant to illustrate two things. The theoretical idea (atomism in this case) does not arise apart from and independent of experience; nor can it be derived from experience by a purely logical procedure. It is produced by a creative act. Once a theoretical idea has been acquired, one does well to hold fast to it until it leads to an untenable conclusion.

AS FOR my latest theoretical work, I do not feel justified in giving a detailed account of it before a wide group of readers interested in science. That should be done only with theories which have been adequately confirmed by experience. So far it is primarily the simplicity of its premises and its intimate connection with what is already known (*viz.*, the laws of the pure gravitational field) that speak in favor of the theory to be discussed here. It may, however, be of interest to a wide group of readers to become acquainted with the train of thought which can lead to endeavors of such an extremely speculative nature. Moreover, it will be shown what kinds of difficulties are encountered and in what sense they have been overcome.

In Newtonian physics the elementary theoretical concept on which the theoretical description of material bodies is based is the material point, or particle. Thus matter is considered *a priori* to be discontinuous. This makes it necessary to consider the action of material points on one another as "action at a distance." Since the latter concept seems quite contrary to everyday experience, it is only natural that the contemporaries of Newton—and indeed Newton himself—found it difficult to accept. Owing to the almost miraculous success of the Newtonian system, however, the succeeding generations of physicists became used to the idea of action at a distance. Any doubt was buried for a long time to come.

But when, in the second half of the 19th century, the laws of electrodynamics became known, it turned out that these laws could not be satisfactorily incorporated into the Newtonian system. It is fascinating to muse: Would Faraday have discovered the law of electromagnetic induction if he had received a regular college education? Unencumbered by the traditional way of thinking, he felt that the introduction of the "field" as an independent element of reality helped him to coordinate the experimental facts. It was Maxwell who fully comprehended the significance of the field concept; he made the fundamental discovery that the laws of electrodynamics found their natural expression in the differential equations for the electric and magnetic fields. These equations implied the existence of waves, whose properties corresponded to those of light as far as they were known at that time.

This incorporation of optics into the theory of electromagnetism represents one of the greatest triumphs in the striving toward unification of the foundations of physics; Maxwell achieved this unification by purely theoretical arguments, long before it was corroborated by Hertz' experimental work. The new insight made it possible to dispense with the hypothesis of action at a distance, at least in the realm of electromagnetic phenomena; the intermediary field now appeared as the only carrier of electromagnetic interaction between bodies, and the field's behavior was completely determined by contiguous processes, expressed by differential equations.

Now a question arose: Since the field exists even in a vacuum, should one conceive of the field as a state of a "carrier," or should it rather be endowed with an independent existence not reducible to anything else? In other words, is there an "ether" which carries the field; the ether being considered in the undulatory state, for example, when it carries light waves?

The question has a natural answer: Because one cannot dispense with the field concept, it is preferable not to introduce in addition a carrier with hypothetical properties. However, the pathfinders who first recognized the indispensability of the field concept were still too strongly imbued with the mechanistic tradition of thought to accept unhesitatingly this simple point of view. But in the course of the following decades this view imperceptibly took hold.

The introduction of the field as an elementary concept gave rise to an inconsistency of the theory as a whole. Maxwell's theory, although adequately describing the behavior of electrically charged particles in their interaction with one another, does not explain the behavior of electrical densities, *i.e.*, it does not provide a theory of the particles themselves. They must therefore be treated as mass points on the basis of the old theory. The combination of the idea of a continuous field with that of material points discontinuous in space appears inconsistent. A consistent field theory requires continuity of all elements of the theory, not only in time but also in space, and in all points of space. Hence the material particle has no place as a fundamental concept in a field theory. Thus even apart from the fact that gravitation is not included, Maxwell's electrodynamics cannot be considered a complete theory.

Maxwell's equations for empty space remain unchanged if the spatial coordinates and the time are subjected to a particular kind of linear transformations—the Lorentz transformations ("covariance" with respect to Lorentz transformations). Covariance also holds, of course, for a transformation which is composed of two or more such transformations; this is called the "group" property of Lorentz transformations.

Maxwell's equations imply the "Lorentz group," but the Lorentz group does not imply Maxwell's equations. The Lorentz group may indeed be defined independently of Maxwell's equations as a group of linear transformations which leave a particular value of the velocity—the velocity of light—invariant. These transformations hold for the transition from one "inertial system" to another which is in uniform motion relative to the first. The most conspicuous novel property of this transformation group is that it does away with the absolute character of the concept of simultaneity of events distant from each other in space. On this account it is to be expected that all equations of physics are covariant with respect to Lorentz transformations (special theory of relativity). Thus it came about that Maxwell's equations led to a heuristic principle valid far beyond the range of the applicability or even validity of the equations themselves.

Special relativity has this in common with Newtonian mechanics: The laws of both theories are supposed to hold only with respect to certain coordinate systems: those known as "inertial systems." An inertial system is a system in a state of motion such that "force-free" material points within it are not accelerated with respect to the coordinate system. However, this definition is empty if there is no independent means for recognizing the absence of forces. But such a means of recognition does not exist if gravitation is considered as a "field."

Let A be a system uniformly accelerated with respect to an "inertial system" I. Material points, not accelerated with respect to I, are accelerated with respect to A, the acceleration of all the points being equal in magnitude and direction. They behave as if a gravitational field exists with respect to A, for it is a characteristic property of the gravitational field that the acceleration is independent of the particular nature of the body. There is no reason to exclude the possibility of interpreting this behavior as the effect of a "true" gravitational field (*principle of equivalence*). This interpretation implies that A is an "inertial system," even though it is accelerated with respect to another inertial system. (It is essential for this argument that the introduction of independent gravitational fields is considered justified even though no masses generating the field are defined. Therefore, to Newton such an argument would not have appeared convincing.) Thus the concepts of inertial system, the law of inertia and the law of motion are deprived of their concrete meaning—not only in classical mechanics but also in special relativity. Moreover, following up this train of thought, it turns out that with respect to A time cannot be measured by identical clocks; indeed, even the immediate physical signi-

ficance of coordinate differences is generally lost In view of all these difficulties, should one not try, after all, to hold on to the concept of the inertial system, relinquishing the attempt to explain the fundamental character of the gravitational phenomena which manifest themselves in the Newtonian system as the equivalence of inert and gravitational mass? Those who trust in the comprehensibility of nature must answer: No.

THIS is the gist of the principle of equivalence: In order to account for the equality of inert and gravitational mass within the theory it is necessary to admit non-linear transformations of the four coordinates. That is, the group of Lorentz transformations and hence the set of the "permissible" coordinate systems has to be extended.

What group of coordinate transformations can then be substituted for the group of Lorentz transformations? Mathematics suggests an answer which is based on the fundamental investigations of Gauss and Riemann: namely, that the appropriate substitute is the group of all continuous (analytical) transformations of the coordinates. Under these transformations the only thing that remains invariant is the fact that neighboring points have nearly the same coordinates; the coordinate system expresses only the topological order of the points in space (including its four-dimensional character). The equations expressing the laws of nature must be covariant with respect to all continuous transformations of the coordinates. This is the principle of general relativity.

The procedure just described overcomes a deficiency in the foundations of mechanics which had already been noticed by Newton and was criticized by Leibnitz and, two centuries later, by Mach: Inertia resists acceleration, but acceleration relative to what? Within the frame of classical mechanics the only answer is: Inertia resists acceleration *relative to space*. This is a physical property of space—space acts on objects, but objects do not act on space. Such is probably the deeper meaning of Newton's assertion *spatium est absolutum* (space is absolute). But the idea disturbed some, in particular Leibnitz, who did not ascribe an independent existence to space but considered it merely a property of "things" (contiguity of physical objects). Had his justified doubts won out at that time, it hardly would have been a boon to physics, for the empirical and theoretical foundations necessary to follow up his idea were not available in the 17th century.

According to general relativity, the concept of space detached from any physical content does not exist. The physical reality of space is represented by a field whose components are continuous functions of four independent variables —the coordinates of space and time. It is just this particular kind of dependence that expresses the spatial character of physical reality

Since the theory of general relativity implies the representation of physical reality by a *continuous* field, the concept of particles or material points cannot play a fundamental part, nor can the concept of motion. The particle can only appear as a limited region in space in which the field strength or the energy density are particularly high.

A relativistic theory has to answer two questions: 1) What is the mathematical character of the field? 2) What equations hold for this field?

Concerning the first question: From the mathematical point of view the field is essentially characterized by the way its components transform if a coordinate transformation is applied. Concerning the second question: The equations must determine the field *to a sufficient extent* while satisfying the postulates of general relativity. Whether or not this requirement can be satisfied depends on the choice of the field-type.

The attempt to comprehend the correlations among the empirical data on the basis of such a highly abstract program may at first appear almost hopeless. The procedure amounts, in fact, to putting the question: What most simple property can be required from what most simple object (field) while preserving the principle of general relativity? Viewed from the standpoint of formal logic, the dual character of the question appears calamitous, quite apart from the vagueness of the concept "simple." Moreover, from the standpoint of physics there is nothing to warrant the assumption that a theory which is "logically simple" should also be "true."

Yet every theory is speculative. When the basic concepts of a theory are comparatively "close to experience" (*e.g.*, the concepts of force, pressure, mass), its speculative character is not so easily discernible. If, however, a theory is such as to require the application of complicated logical processes in order to reach conclusions from the premises that can be confronted with observation, everybody becomes conscious of the speculative nature of the theory. In such a case an almost irresistible feeling of aversion arises in people who are inexperienced in epistemological analysis and who are unaware of the precarious nature of theoretical thinking in those fields with which they are familiar.

On the other hand, it must be conceded that a theory has an important advantage if its basic concepts and fundamental hypotheses are "close to experience," and greater confidence in such a theory is certainly justified. There is less danger of going completely astray, particularly since it takes so much less time and effort to disprove such theories by experience. Yet more and more, as the depth of our knowledge increases, we must give up this advantage in our quest for logical simplicity and uniformity in the foundations of physical theory. It has to be admitted that general relativity has gone further than previous physical theories in relinquishing "closeness to experience" of fundamental concepts in order to attain logical simplicity. This holds already for the theory of gravitation, and it is even more true of the new generalization, which is an attempt to comprise the properties of the total field. In the generalized theory the procedure of deriving from the premises of the theory conclusions that can be confronted with empirical data is so difficult that so far no such result has been obtained. In favor of this theory are, at this point, its logical simplicity and its "rigidity." Rigidity means here that the theory is either true or false, but not modifiable.

THE greatest inner difficulty impeding the development of the theory of relativity is the dual nature of the problem, indicated by the two questions we have asked. This duality is the reason why the development of the theory has taken place in two steps so widely separated in time. The first of these steps, the theory of gravitation, is based on the principle of equivalence discussed above and rests on the following consideration: According to the theory of special relativity, light has a constant velocity of propagation. If a light ray in a vacuum starts from a point, designated by the coordinates x_1, x_2 and x_3 in a three dimensional coordinate system, at the time x_4, it spreads as a spherical wave and reaches a neighboring point (x_1+dx_1, x_2+dx_2, x_3+dx_3) at the time x_4+dx_4. Introducing the velocity of light, c, we write the expression:

$$\sqrt{dx_1{}^2+dx_2{}^2+dx_3{}^2}=cdx_4$$

This can also be written in the form:

$$dx_1{}^2+dx_2{}^2+dx_3{}^2-c^2\,dx_4{}^2=0$$

This expression represents an objective relation between neighboring space-time points in four dimensions, and it holds for all inertial systems, provided the coordinate transformations are restricted to those of special relativity. The relation loses this form, however, if arbitrary continuous transformations of the coordinates are admitted in accordance with the principle of general relativity. The relation then assumes the more general form:

$$\sum_{ik} g_{ik}\,dx_i\,dx_k=0$$

The g_{ik} are certain functions of the coor-

dinates which transform in a definite way if a continuous coordinate transformation is applied. According to the principle of equivalence, these g_{ik} functions describe a particular kind of gravitational field: a field which can be obtained by transformation of "field-free" space. The g_{ik} satisfy a particular law of transformation. Mathematically speaking, they are the components of a "tensor" with a property of symmetry which is preserved in all transformations; the symmetrical property is expressed as follows:

$$g_{ik}=g_{ki}$$

The idea suggests itself: May we not ascribe objective meaning to such a symmetrical tensor, even though the field *cannot* be obtained from the empty space of special relativity by a mere coordinate transformation? Although we cannot expect that such a symmetrical tensor will describe the most general field, it may well describe the particular case of the "pure gravitational field." Thus it is evident what kind of field, at least for a special case, general relativity has to postulate: a symmetrical tensor field.

Hence only the second question is left: What kind of general covariant field law can be postulated for a symmetrical tensor field?

This question has not been difficult to answer in our time, since the necessary mathematical conceptions were already at hand in the form of the metric theory of surfaces, created a century ago by Gauss and extended by Riemann to manifolds of an arbitrary number of dimensions. The result of this purely formal investigation has been amazing in many respects. The differential equations which can be postulated as field law for g_{ik} cannot be of lower than second order, *i.e.*, they must at least contain the second derivatives of the g_{ik} with respect to the coordinates. Assuming that no higher than second derivatives appear in the field law, *it is mathematically determined by the principle of general relativity*. The system of equations can be written in the form:

$$R_{ik}=0$$

The R_{ik} transform in the same manner as the g_{ik}, *i.e.*, they too form a symmetrical tensor.

These differential equations completely replace the Newtonian theory of the motion of celestial bodies provided the masses are represented as singularities of the field. In other words, they contain the law of force as well as the law of motion while eliminating "inertial systems."

The fact that the masses appear as singularities indicates that these masses themselves cannot be explained by symmetrical g_{ik} fields, or "gravitational fields." Not even the fact that only *positive* gravitating masses exist can be deduced from this theory. Evidently a complete relativistic field theory must be based on a field of more complex nature, that is, a generalization of the symmetrical tensor field.

BEFORE considering such a generalization, two remarks pertaining to gravitational theory are essential for the explanation to follow.

The first observation is that the principle of general relativity imposes exceedingly strong restrictions on the theoretical possibilities. Without this restrictive principle it would be practically impossible for anybody to hit on the gravitational equations, not even by using the principle of special relativity, even though one knows that the field has to be described by a symmetrical tensor. No amount of collection of facts could lead to these equations unless the principle of general relativity were used. This is the reason why all attempts to obtain a deeper knowledge of the foundations of physics seem doomed to me unless the basic concepts are in accordance with general relativity from the beginning. This situation makes it difficult to use our empirical knowledge, however comprehensive, in looking for the fundamental concepts and relations of physics, and it forces us to apply free speculation to a much greater extent than is presently assumed by most physicists. I do not see any reason to assume that the heuristic significance of the principle of general relativity is restricted to gravitation and that the rest of physics can be dealt with separately on the basis of special relativity, with the hope that later on the whole may be fitted consistently into a general relativistic scheme. I do not think that such an attitude, although historically understandable, can be objectively justified. The comparative smallness of what we know today as gravitational effects is not a conclusive reason for ignoring the principle of general relativity in theoretical investigations of a fundamental character. In other words, I do not believe that it is justifiable to ask: What would physics look like without gravitation?

The second point we must note is that the equations of gravitation are 10 differential equations for the 10 components of the symmetrical tensor g_{ik}. In the case of a non-general relativistic theory, a system is ordinarily not overdetermined if the number of equations is equal to the number of unknown functions. The manifold of solutions is such that within the general solution a certain number of functions of three variables can be chosen arbitrarily. For a general relativistic theory this cannot be expected as a matter of course. Free choice with respect to the coordinate system implies that out of the 10 functions of a solution, or components of the field, four can be made to assume prescribed values by a suitable choice of the coordinate system. In other words, the principle of general relativity implies that the number of functions to be determined by differential equations is not 10 but $10-4=6$. For these six functions only six independent differential equations may be postulated. Only six out of the 10 differential equations of the gravitational field ought to be independent of each other, while the remaining four must be connected to those six by means of four relations (identities). And indeed there exist among the left-hand sides, R_{ik}, of the 10 gravitational equations four identities—"Bianchi's identities"—which assure their "compatibility."

In a case like this—when the number of field variables is equal to the number of differential equations—compatibility is always assured if the equations can be obtained from a variational principle. This is indeed the case for the gravitational equations.

However, the 10 differential equations cannot be entirely replaced by six. The system of equations is indeed "overdetermined," but due to the existence of the identities it is overdetermined in such a way that its compatibility is not lost, *i.e.*, the manifold of solutions is not critically restricted. The fact that the equations of gravitation imply the law of motion for the masses is intimately connected with this (permissible) overdetermination.

After this preparation it is now easy to understand the nature of the present investigation without entering into the details of its mathematics. The problem is to set up a relativistic theory for the total field. The most important clue to its solution is that there exists already the solution for the special case of the pure gravitational field. The theory we are looking for must therefore be a generalization of the theory of the gravitational field. The first question is: What is the natural generalization of the symmetrical tensor field?

This question cannot be answered by itself, but only in connection with the other question: What generalization of the field is going to provide the most natural theoretical system? The answer on which the theory under discussion is based is that the symmetrical tensor field must be replaced by a non-symmetrical one. This means that the condition $g_{ik}=g_{ki}$ for the field components must be dropped. In that case the field has 16 instead of 10 independent components.

There remains the task of setting up the relativistic differential equations for a non-symmetrical tensor field. In the attempt to solve this problem one meets with a difficulty which does not arise in the case of the symmetrical field. The principle of general relativity does not suffice to determine completely the field equations, mainly because the transformation law of the symmetrical part of the field alone does not involve the components of the antisymmetrical part or

vice versa. Probably this is the reason why this kind of generalization of the field has hardly ever been tried before. The combination of the two parts of the field can only be shown to be a natural procedure if in the formalism of the theory only the total field plays a role, and not the symmetrical and antisymmetrical parts separately.

It turned out that this requirement can indeed be satisfied in a natural way. But even this requirement, together with the principle of general relativity, is still not sufficient to determine uniquely the field equations. Let us remember that the system of equations must satisfy a further condition: the equations must be compatible. It has been mentioned above that this condition is satisfied if the equations can be derived from a variational principle.

This has indeed been achieved, although not in so natural a way as in the case of the symmetrical field. It has been disturbing to find that it can be achieved in two different ways. These variational principles furnished two systems of equations—let us denote them by E_1 and E_2—which were different from each other (although only slightly so), each of them exhibiting specific imperfections. Consequently even the condition of compatibility was insufficient to determine the system of equations uniquely.

It was, in fact, the formal defects of the systems E_1 and E_2 that indicated a possible way out. There exists a third system of equations, E_3, which is free of the formal defects of the systems E_1 and E_2 and represents a combination of them in the sense that every solution of E_3 is a solution of E_1 as well as of E_2. This suggests that E_3 may be the system we have been looking for. Why not postulate E_3, then, as the system of equations? Such a procedure is not justified without further analysis, since the compatibility of E_1 and that of E_2 do not imply compatibility of the stronger system E_3, where the number of equations exceeds the number of field components by four.

An independent consideration shows that irrespective of the question of compatibility the stronger system, E_3, is the only really natural generalization of the equations of gravitation.

But E_3 is not a compatible system in the same sense as are the systems E_1 and E_2, whose compatibility is assured by a sufficient number of identities, which means that every field that satisfies the equations for a definite value of the time has a continuous extension representing a solution in four-dimensional space. The system E_3, however, is not extensible in the same way. Using the language of classical mechanics we might say: In the case of the system E_3 the "initial condition" cannot be freely chosen. What really matters is the answer to the question: Is the manifold of solutions for the system E_3 as extensive as must be required for a physical theory? This purely mathematical problem is as yet unsolved.

The skeptic will say: "It may well be true that this system of equations is reasonable from a logical standpoint. But this does not prove that it corresponds to nature." You are right, dear skeptic. Experience alone can decide on truth. Yet we have achieved something if we have succeeded in formulating a meaningful and precise question. Affirmation or refutation will not be easy, in spite of an abundance of known empirical facts. The derivation, from the equations, of conclusions which can be confronted with experience will require painstaking efforts and probably new mathematical methods.

The Author

Wherever we look, the physics of the 20th century bears the indelible imprint of the genius of Albert Einstein. His "photoelectric effect," propounded in 1905, set the cornerstone of quantum theory. At several decisive times in the development of that theory it was Albert Einstein who supplied the ideas that ensured its arrival in the dominant role it plays in physics today.

In the same *annus mirabilis* of 1905 Einstein also published two papers that launched a parallel revolution in physical thought—the theory of relativity with which his name is primarily identified. The second of these two papers set forth his celebrated deduction of the equivalence of mass and energy: $E=mc^2$. By 1917 he had built upon his "special theory" of relativity the great edifice of his "general theory" that subsumes the large-scale mechanics of the universe into a comprehensive space-time geometry. From that time on he set himself the lofty aim of bringing into this grand generalization the electromagnetic laws that govern the small-scale realm of atomic particles. In his quest he found himself increasingly alone over the last four decades of his life, out of sympathy with the philosophic views that had come from the quantum theory to hold sway among his contemporaries in physics.

In the present article Einstein undertakes to explain to the layman his last and still unsuccessful effort to formulate a "Generalized Theory of Gravitation." He was himself the most engaging and successful populariser of his work, but he described this article as "not quite easy to grasp." Nonetheless, it offers the reader a warm and intimate insight into the motives and aims of this great natural philosopher.

Bibliography

THE MEANING OF RELATIVITY. Albert Einstein. Princeton University Press, 1950.

RELATIVITY. Albert Einstein. Peter Smith, 1931

THE HYDROGEN BOMB: II

ATOMIC ENERGY. George Gamow. Cambridge University Press, 1947.

ELEMENTARY NUCLEAR THEORY. H. A. Bethe. John Wiley & Sons, Inc., 1947.

SCIENTIFIC AMERICAN September 1956, Vol. 195, No. 3, pp. 82-91 OFFPRINT 210

THE ORIGIN OF THE ELEMENTS

by William A. Fowler

The relative abundance of the various kinds of atoms is a powerful clue to the history of the universe. A discussion of the recent theory that the heavier ones were built up from hydrogen in stars.

In investigating the nature and history of the universe we can hardly do better than to begin by examining what it is made of. The universe we see and measure is composed of an orderly yet diverse system of elements, from hydrogen to uranium. How did these elements come into being; from what primordial stuff were they made? As rare Ben Jonson shrewdly observed more than 300 years ago in *The Alchemist* (in a quotation to which the physicists Ralph A. Alpher and Robert C. Herman have previously called attention):

Ay, for 'twere absurd
To think that nature in the earth
bred gold
Perfect i' the instant: something
went before.
There must be remote matter.

Research into "remote matter" and the origin of the elements is going forward today along many paths, and of these none has been more fruitful than the study of the relative abundance of the various elements in the universe. The present abundances of the elements offer one of our most powerful clues to the history of the earth, the stars and the galaxies, for the abundance curve is the product of that history and was shaped by cosmic events. From this curve we can learn much about the evolution of stars, about cosmology and about all the grand-scale subjects of modern science.

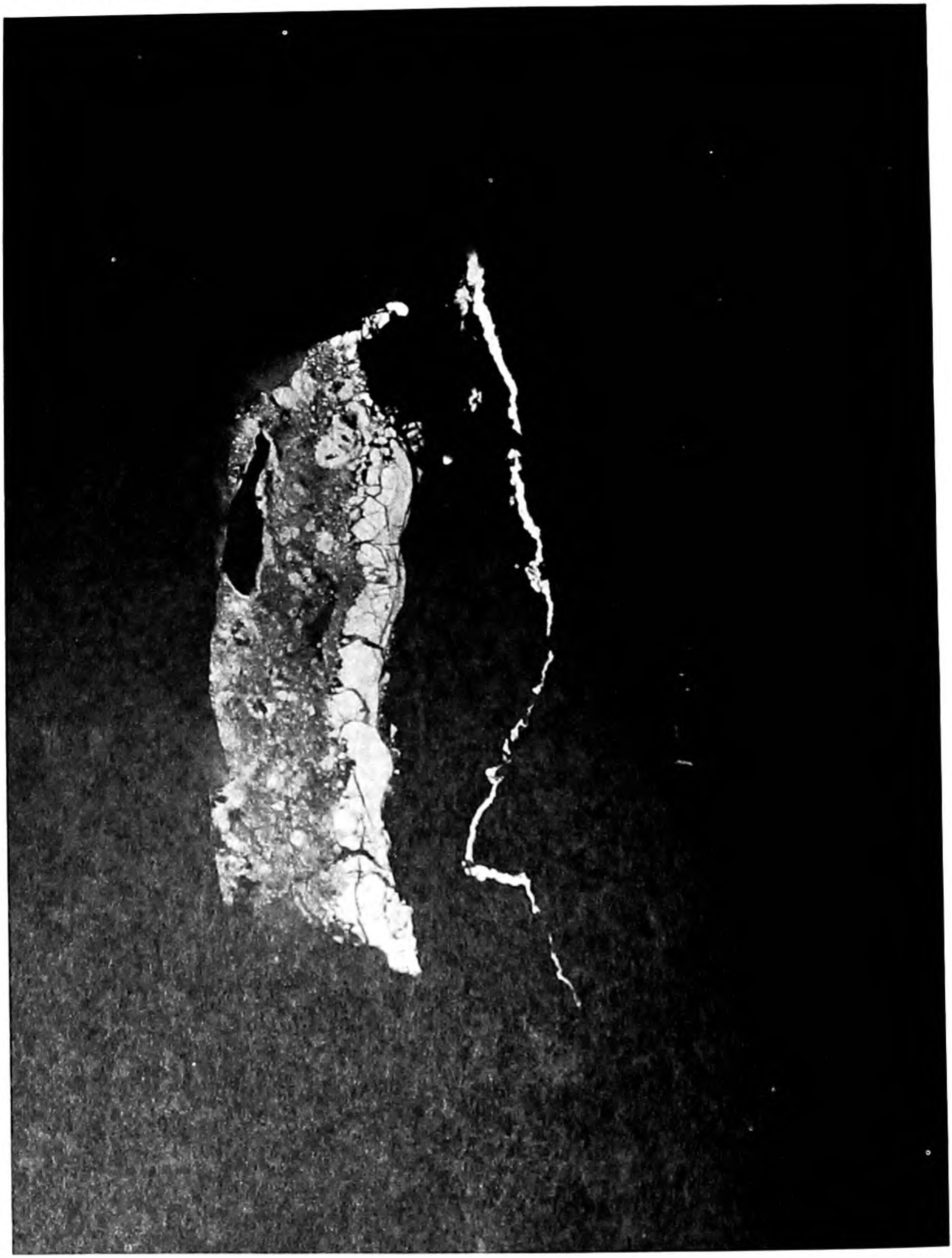
Our inquiries into the composition of the universe are severely handicapped, to be sure, by the fact that gravitation, which acts alike on heavenly bodies, apples and human beings, has so far chained mankind to his native planet. But notwithstanding this handicap, an imposing range of information on the universal abundance of elements is available to us today. There is, first of all, our own planet, where we can analyze at first hand the composition of the crust, oceans and atmosphere, and, allowing for losses of matter to space and redistribution of matter to the interior, can compute the proportions of the elements in the earth when it was formed. Secondly, there are the meteorites plucked by the earth from outer space; we attach considerable weight to these samples, because the matter in meteorites is assumed to have undergone less change than that in the earth's crust. Thirdly, the light from a star, when analyzed with the spectroscope, identifies the elements on its visible surface. Every element emits or absorbs a characteristic spectrum of light (bright or dark lines at certain wavelengths) when its atoms are excited to high temperature; the elements have been "fingerprinted" in this way in laboratories, and their prints can be matched to the spectral light from stars. The abundance of each element can be estimated from the intensity of its radiation or from the amount of radiation the surface atoms absorb from the star's background radiation. Fourthly, from galaxies and from interstellar space we can hear a song of hydrogen, in the form of radio waves at the 21-centimeter wavelength; as radio astronomy develops it may tell us much more about the abundance of the elements in space. Finally, the cosmic ray particles that continually bombard the earth also supply us with samples of matter from the universe outside our planet.

All these clues are beset with complications that may mislead us. Nor can we be confident that we have a true sample of the whole universe, for the information comes mainly from our own galaxy, indeed, largely from our own solar system. But it has been gratifying to find that every one of our methods of observation, when carefully carried out and corrected for complicating factors, yields much the same story. They produce a reasonable and consistent picture of the average abundance of the elements in the universe as far as we can observe it. This picture—a curve showing the proportions of the various elements in the cosmos as a whole—is well represented by the curve constructed by Harrison Brown of the California Institute of Technology on the basis of his analysis of meteorites and other evidence [*see upper chart on page 69*].

By far the most abundant element is hydrogen: it accounts for 93 per cent of the total number of atoms and 76 per cent of the weight of the universe's matter. Helium is next: about 7 per cent by number of atoms and 23 per cent by weight. In general the abundance of the elements drops off with increasing atomic weight. The fall in the curve has one sharp interruption when we come to the elements of the iron group: these are about 10,000 times more abundant than their neighbors in the atomic-weight sequence. But except for this anomaly there is a general decline, and the heaviest elements add up to only a hundred millionth of all matter by number of atoms and a millionth by weight. It is a striking fact that all the elements beyond helium together amount to only a little more than 1 per cent of the mass of the universe.

If we take this picture to be correct, we have, then, a universal pudding composed of certain known ingredients mixed in certain proportions. Our task is to determine what recipe could have brewed this mixture.

We begin with the fact that, to the best of our knowledge, all the elements



URANIUM-BEARING ROCK, when cut so that it can be placed in close contact with a glass photographic negative, records its radioactivity in the emulsion. The relative abundance of the isotopes in such rocks provides evidence as to the age of the earth.

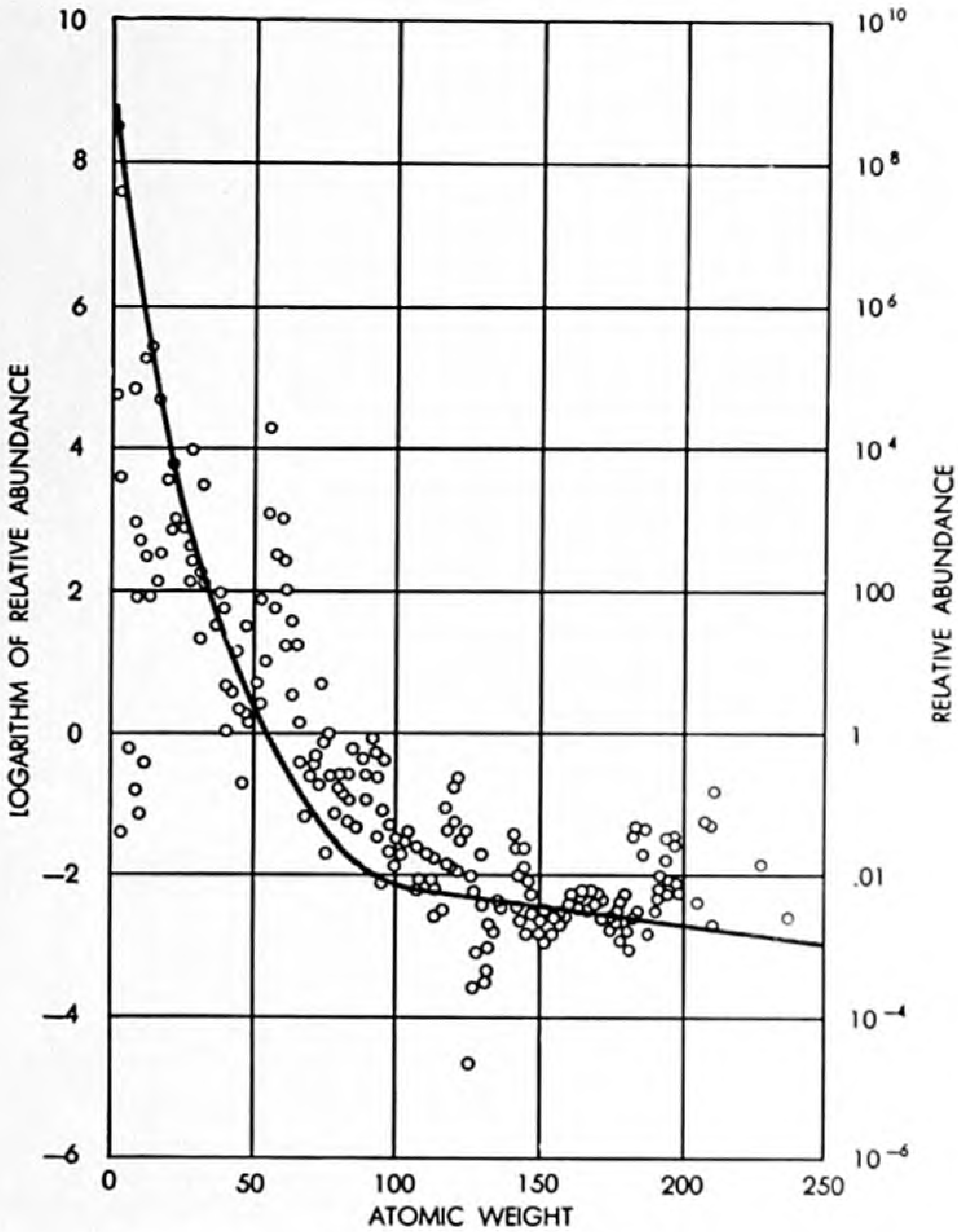
are made up of just two nuclear building blocks—protons and neutrons. (How the protons and neutrons themselves were created is a question outside the province of this article: only men of strong convictions, religious or scientific, have the courage to deal with the problem of the creation.) In a sense protons and neutrons are merely different versions of a nucleon: a free neutron may decay into a proton by shedding a negative electron, and the positively charged proton may become a neutron by combining with an electron or by emitting a positron.

The nucleus of the simplest element, hydrogen, is a single proton. Nearly a century and a half ago the Englishman William Prout suggested that all the elements consisted of combinations of hydrogen atoms. We have learned that the situation is vastly more complicated, but essentially most of the modern theories make a similar approach. It is natural to start with the working hypothesis that the elements were built up from protons or neutrons or both as the units.

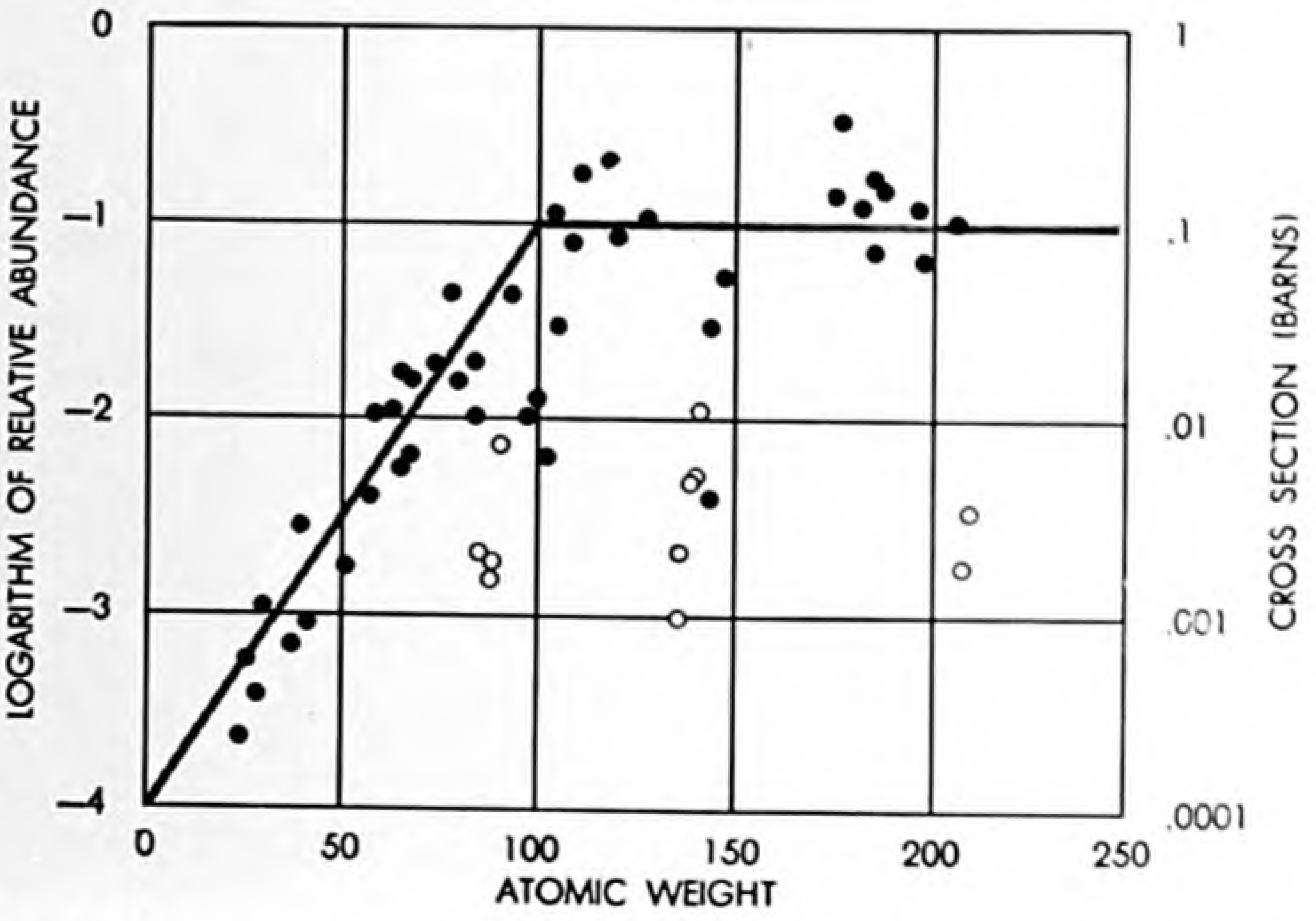
The difficulty lies in trying to picture how this build-up took place and how it could have proceeded through the whole sequence to produce all the elements in the periodic table. The positively charged protons repel each other, and it takes a large amount of energy to overcome this repulsion and force them close enough together to combine. Some combinations are highly unstable or nonexistent. Other combinations in the sequence are so stable and so strongly bound that it is difficult to see how they can be transmuted or built up to larger atoms by natural processes.

There are several current theories about the origin of the elements, but we shall consider only the two that have been worked out in fairly comprehensive fashion and are taken most seriously.

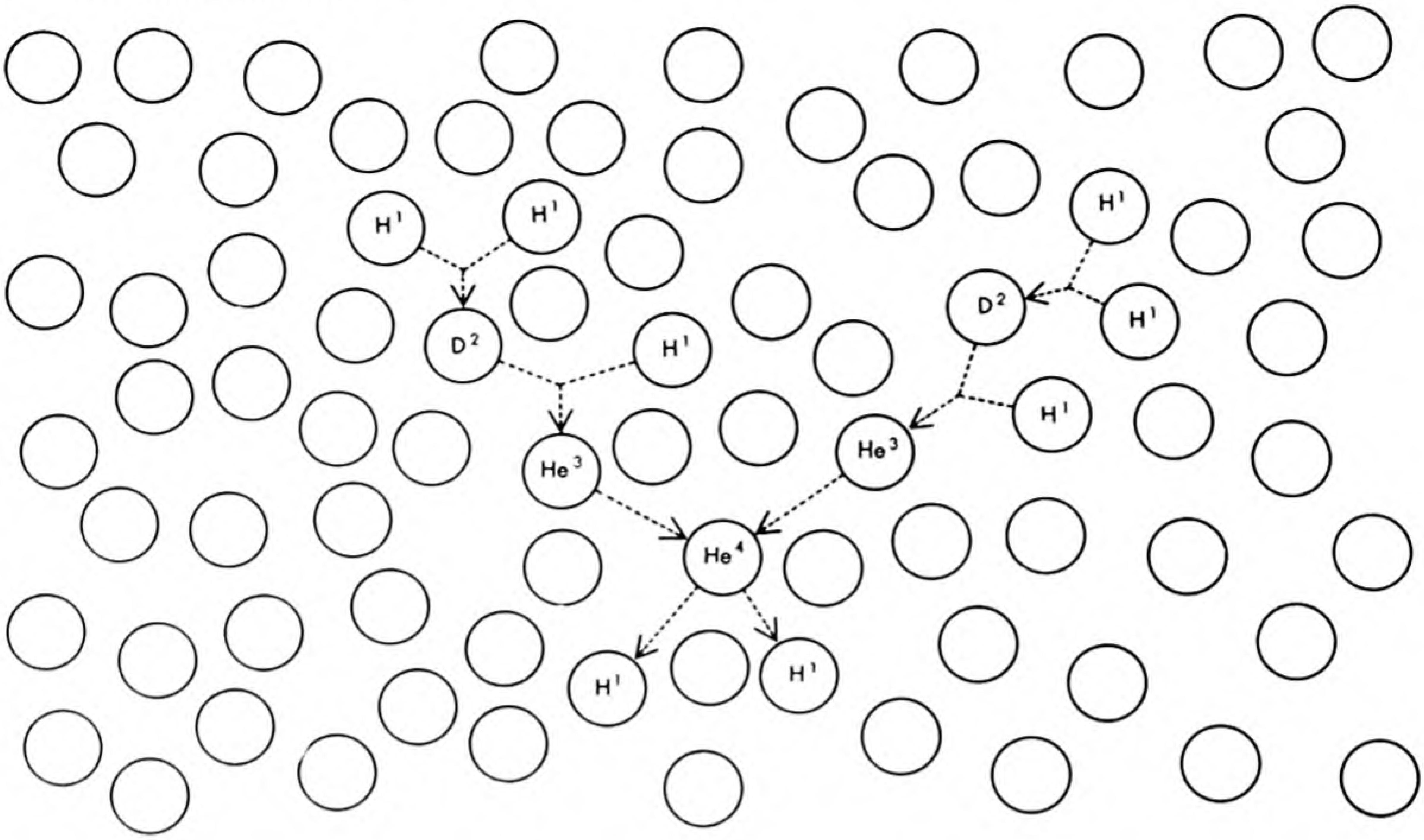
The more popular of the two is the one advanced by George Gamow and his collaborators. This theory holds that the elements were formed by a step-by-step build-up from neutrons. Gamow starts from the postulate, based on the apparent expansion of the universe, that the cosmos started from a core which exploded in a primordial "big bang" some five billion years ago. This exceedingly dense core, he believes, was made up primarily of neutrons, for under the great pressure electrons would be compressed into the protons. As the great neutron ball began to expand, some of the neutrons decayed to protons. Each proton promptly captured a neutron, the



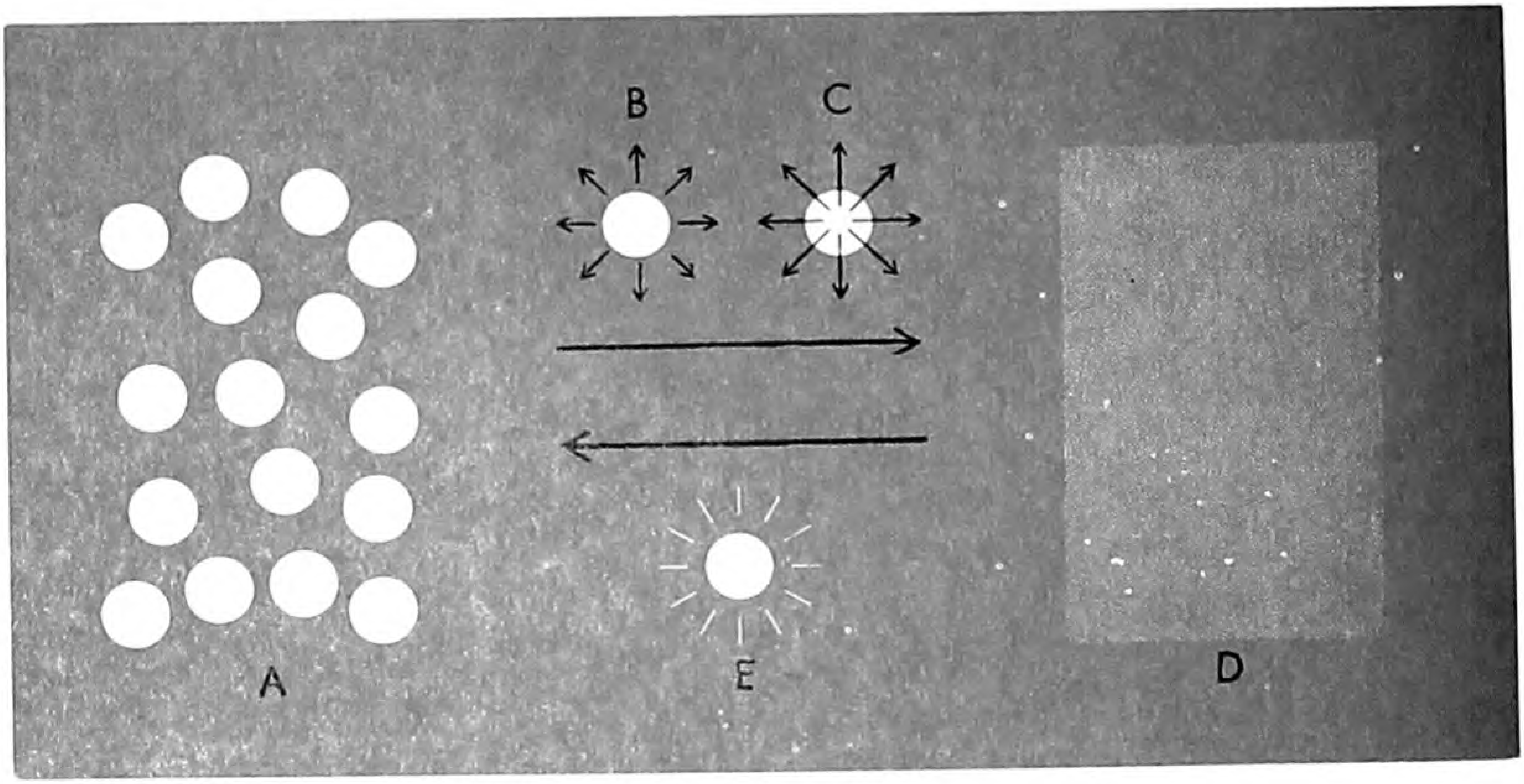
RELATIVE ABUNDANCE of the elements is plotted. The solid curve itself is based on theoretical abundance expected from neutron-capture cross sections (*see below*). Anomalies near atomic weight 56 are the iron group; near 10, rare lithium, beryllium and boron.



CROSS SECTIONS of nuclei for neutron capture are plotted according to Donald J. Hughes and his collaborators. The heavier nuclei capture neutrons more readily. The exceptionally stable nuclei are indicated by open circles. A barn (*type at right*) is 10^{-24} square centimeters.



SYNTHESIS OF ELEMENTS from hydrogen begins with the steps depicted here. The circles represent nuclei in the interior of a star. Two nuclei of hydrogen 1 (H^1) fuse to form a nucleus of hydrogen 2, or deuterium (D^2). The deuterium then fuses with hydrogen 1 to form helium 3 (He^3). Two nuclei of helium 3 next combine to form helium 4 (He^4) and two nuclei of hydrogen 1.

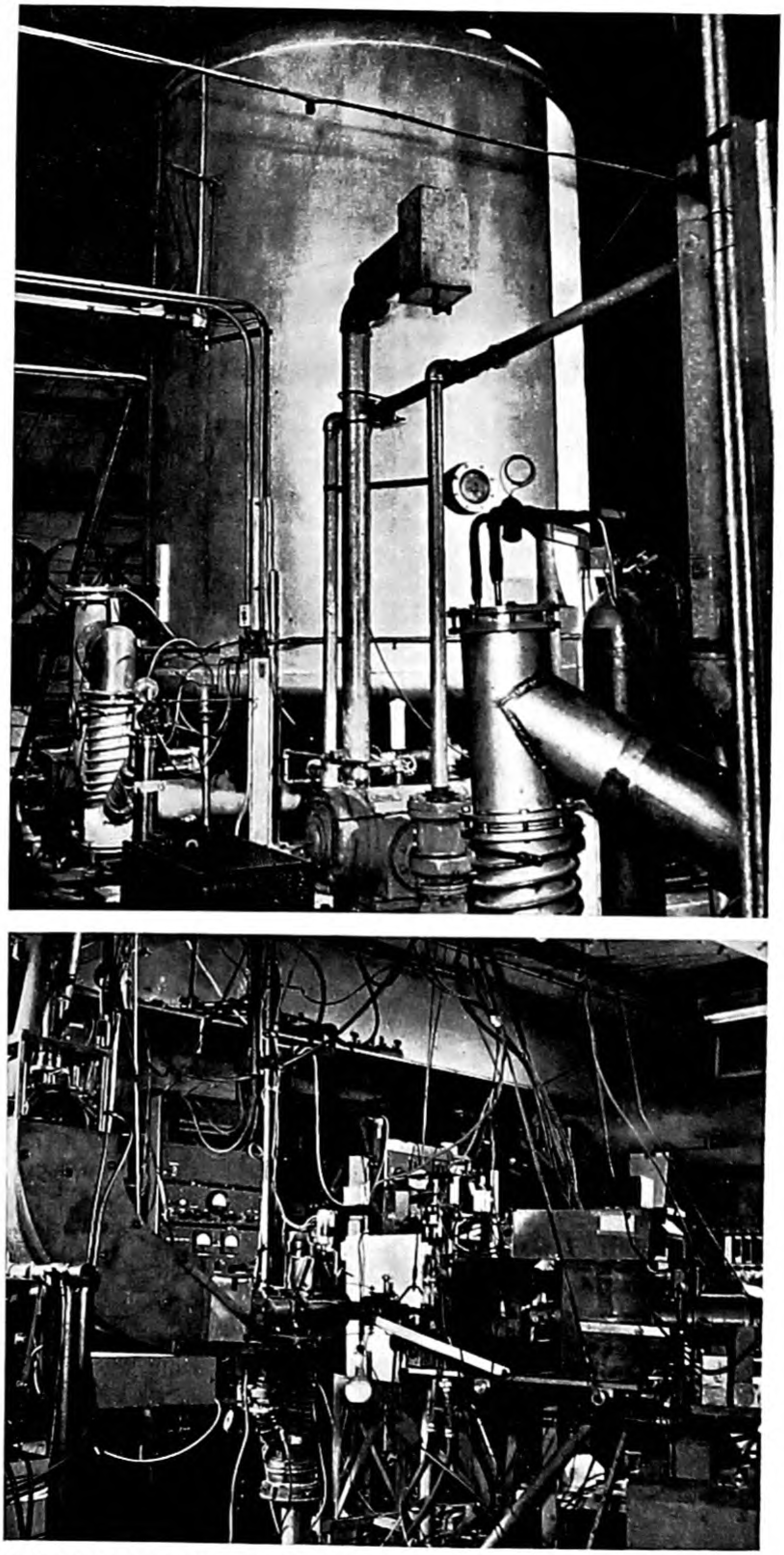


TRANSFER OF ELEMENTS between stars and the interstellar dust and gas is illustrated. Elements are synthesized in stars (A) by processes of the kind depicted at the top of the page. The stellar material is ejected slowly by stars such as the sun (B), or explosively by novae or supernovae (C). The material is mixed in the dust and gas (D). It condenses into young, bright stars (E).

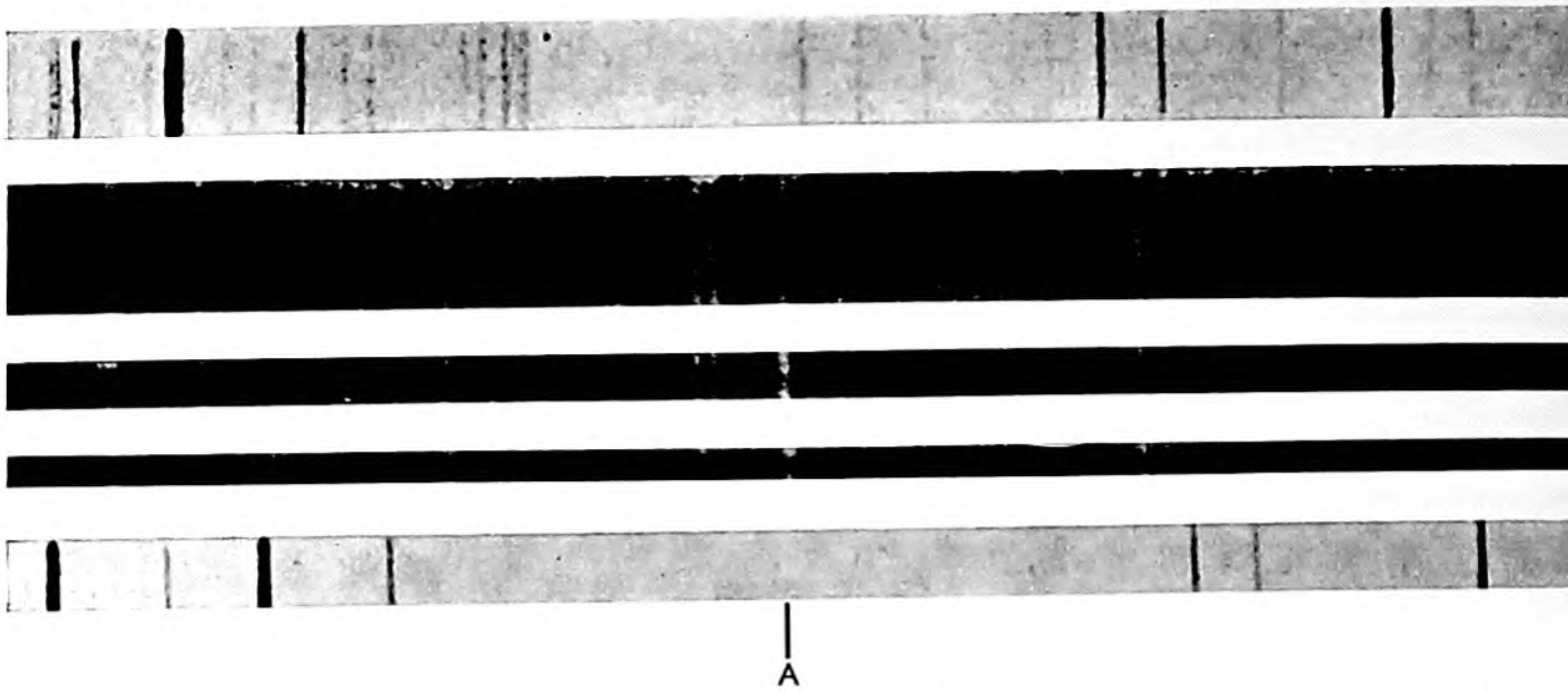
pair forming a deuteron, the nucleus of the hydrogen isotope of mass 2. Some deuterons then captured another neutron and became nuclei of tritium, or hydrogen 3. This nucleus soon decays by emitting a negative electron and thus is transmuted to helium 3. And so, by a rapid succession of neutron captures and electron decays, all the elements were built in the first burst of the universe's expansion. Gamow believes that the whole process of formation of the elements as we know them took place in a matter of a few minutes. The fleeing matter thereafter formed stars, planets and galaxies.

Two lines of evidence from laboratory experiments with particles give impressive support to Gamow's theory. First, it is well established that nearly all nuclei do in fact capture neutrons readily. Secondly, the neutron-capture cross sections of the various nuclei predict a pattern of element abundances which agrees remarkably well with the one actually observed. We should expect a simple relation between the neutron cross section of a given nucleus (*i.e.*, the rate at which it captures neutrons) and the relative abundance of its production. Nuclei that capture neutrons rapidly should be comparatively rare when the sequence of element formation is completed, because most of them are quickly converted by such capture to other nuclei; conversely, nuclei that are slow to capture neutrons should accumulate to relatively high abundance. The curve of element abundances does in fact closely follow the curve of neutron-capture cross sections, in an inverse sense: that is to say, just as the curve of abundance falls sharply from hydrogen to the nucleus of atomic weight 100 and then flattens out, so the curve of neutron cross sections rises sharply from hydrogen to 100 and similarly flattens out beyond this atomic weight [*compare charts on page 69*]. There are even some correlations between fluctuations of elements from the two curves, notably at the neutron numbers 50, 82 and 126.

But there are important difficulties with Gamow's theory—difficulties to which his collaborators Ralph A. Alpher and Robert C. Herman have themselves called attention. The most serious is the fact that in the sequence of atomic weights numbers 5 and 8 are vacant. That is, there is no stable atom of mass 5 or of mass 8. We can produce helium 5 in the laboratory by bombarding helium 4 with neutrons, but it immediately breaks down to helium 4 again. Likewise we can produce momentarily an



ELECTROSTATIC PARTICLE ACCELERATOR is used at the California Institute of Technology to study nuclear reactions of the kind that occur in stars. In the photograph at the top is the vacuum tank which encloses the generator. In the photograph at the bottom is the apparatus on the floor below. The beam of particles is bent by curved magnet at left.



STELLAR SPECTRA studied by G. R. and E. Margaret Burbidge show an intensification of the absorption lines due to the elements barium, lanthanum, praseodymium and samarium in two "peculiar" stars. Reproduced here are five spectra. At the top and bottom, for comparison purposes, are the spectra of an iron arc. Second from the top is the spectrum of the normal star Kappa

isotope of beryllium of mass 8, but it too instantly breaks down (by fission into two helium 4 atoms). The question then is: How can the build-up of elements by neutron capture get by these gaps? The process could not go beyond helium 4, and even if it spanned this gap it would be stopped again at mass 8. In short, if neutron capture were the only process by which elements could be built, starting with hydrogen, the build-up would get no farther than helium.

This basic objection to Gamow's theory is a great disappointment, in view of the promise and philosophical attractiveness of the idea. The other major current hypothesis is less simple and less elegant; it complicates the picture by invoking other processes, in addition to neutron capture, to account for the build-up of the elements. But it seems to surmount the difficulties encountered by the Gamow hypothesis.

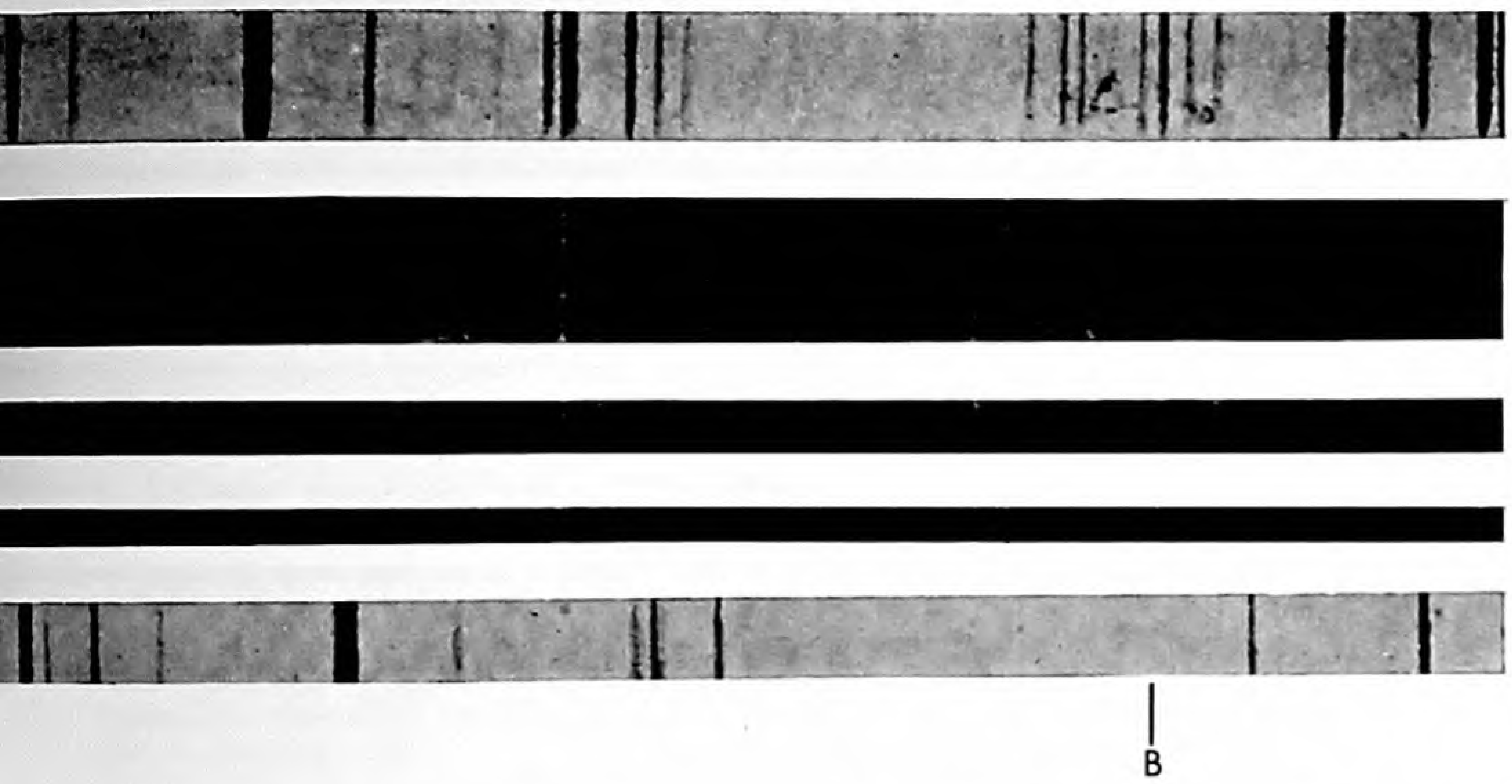
The theory argues that the elements were built not in a primordial explosion but in the hot interiors of stars. It starts from our knowledge that nuclear reactions and transformations must be going on constantly in the stars. As Sir Arthur Eddington presciently remarked in 1920, after Lord Rutherford had transmuted nuclei by bombardment in his laboratory: "What is possible in the Cavendish Laboratory may not be too difficult in the sun." Eddington's informed guess was certainly correct, but not until 1938 was it translated into terms of specific processes. Hans A. Bethe, seeking to account for the enormous and enduring energy of the sun and other stars, conceived two chains of nuclear reactions that would explain their tremendous release of energy and would build new nuclei. The processes have been known ever since as proton-proton fusion and the carbon-nitrogen cycle. The new theory of synthesis of the elements, which has been championed most extensively by Fred Hoyle of the University of Cambridge, assigns key roles to these processes.

We start with a universe consisting of a cold, dilute and turbulent gas of hydrogen atoms. By gravitational attraction part of the gas condenses into stars. As a star contracts under gravitational force, its interior grows very dense and hot. When the central temperature reaches about five million degrees, the protons are moving with enough energy to fuse on colliding and form deuterons. Deuterons in turn combine with protons to form helium 3. Helium 3 does not interact with protons, but laboratory experiments have shown that two helium 3 nuclei can fuse and produce helium 4, ejecting the two surplus protons. The net result of this proton-proton chain is the conversion of four atoms of hydrogen into one atom of helium.

In this way a core of helium develops in the center of the star and gradually grows in size. After a time, as the hydrogen fuel in the interior is used up, the core begins to cool. It then contracts, because gravitational forces gain the upper hand. As a result the temperature of the core rises again. The sudden rise of the internal temperature heats up the star's outer envelope of hydrogen; the mantle expands enormously; its extended surface then radiates cooler (*i.e.*, redder) light, and the star becomes a "red giant."

We have now a star with a hot core of helium, at a computed temperature of more than 100 million degrees. What happens next? We have come to the Gordian knot of the speculations on the build-up of the elements. Two helium nuclei may combine to form a nucleus of mass 8, but as we have seen, any nucleus of mass 8 must be extremely unstable, for none is found in nature. However, beryllium 8 *has* been produced momentarily in the laboratory, and will certainly materialize in the very hot and dense interior of a star. In fact, in that environment beryllium 8 will be produced at as fast a rate as it breaks down, so that a small amount of it is always present. If so, an occasional beryllium 8 nucleus may during its very brief lifetime fuse with a helium 4 nucleus. The combination should result in a nucleus of carbon 12.

Hoyle has pointed out that, in view of the extreme rarity of the beryllium 8 nuclei (about one part in 10 billion in a



Geminorum. Third from the top is the spectrum of the peculiar star HD 46407; fourth from the top, the spectrum of the peculiar star HD 26. At A is an absorption line for ionized barium; at B, an absorption line for ionized lanthanum. The intensity of these lines in the spectra of the peculiar stars shows that their atmospheres contain more barium and lanthanum than that of Kappa Geminorum.

100-million-degree star), the beryllium 8 nucleus had better have a big cross section for capturing helium nuclei if this scheme is to work. Naturally the question cannot be put to a direct test by bombarding a beryllium 8 target in the laboratory, for the nucleus is too ephemeral. But in the W. K. Kellogg Radiation Laboratory at Cal Tech we have been able to obtain indirect evidence that this capture does have a high probability, or, in the parlance of nuclear physics, that it is a "resonant" reaction. Hoyle reasoned that if the reaction is indeed a resonant one, the product, carbon 12, must go through an excited state with certain specified properties. We have found that the carbon 12 nucleus can in fact take this excited form, with almost exactly the properties Hoyle predicted. We produced excited carbon by bombarding boron with high-energy deuterons. The excited carbon 12 nucleus resulting from this reaction promptly disintegrated into three helium nuclei. On the basis of very general physical principles we can argue that in the hot core of helium in a star the reverse process can take place: that is, three helium nuclei may combine to form excited carbon 12, which may then discharge its energy of excitation and become stable carbon.

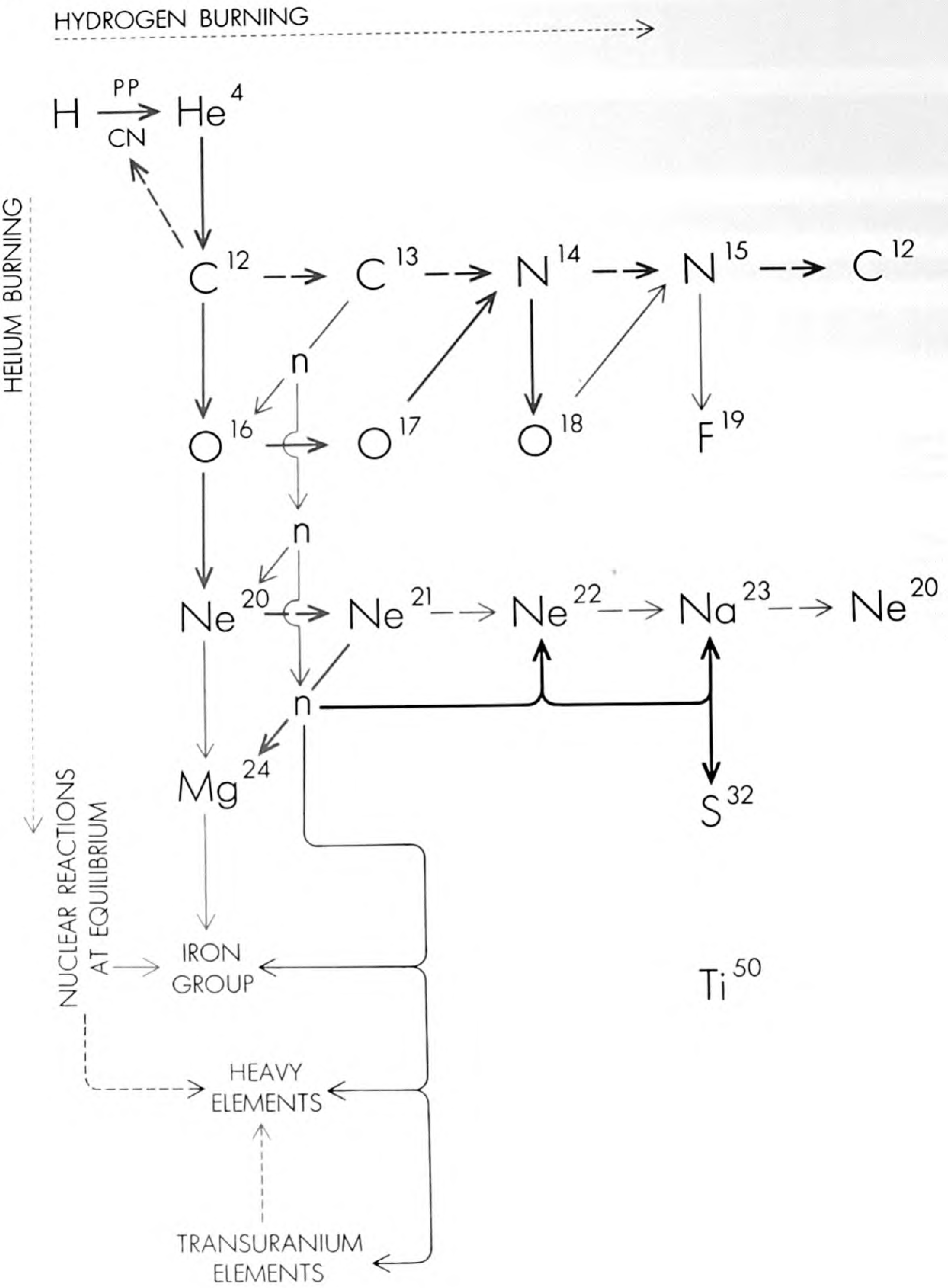
The jump from helium to carbon of course skips the elements lithium, beryllium (whose stable form is beryllium 9) and boron. There is good reason to suppose that these elements are not produced in the main line of build-up of the elements. They are comparatively rare, and may be made by secondary processes. It is known, for instance, that bombardment of heavy elements with hydrogen nuclei sometimes chips off fragments which are identifiable as nuclei of lithium, beryllium and boron. Possibly this process goes on in spots ("sunspots") on the surfaces of stars or occurs in stellar explosions.

Once carbon 12 has been synthesized in the helium core of a star, it may build up by successive captures of helium nuclei to oxygen 16, neon 20 and perhaps magnesium 24. When the helium has been largely used up, so that there can no longer be much release of energy from these fusion reactions, the core cools and contracts. The contraction again raises the temperature of the core, this time perhaps to an energy high enough to trigger interactions among the nuclei of carbon, oxygen and neon. Such reactions would produce the silicon group of elements (around atomic weight 28). The temperature of the core may continue to rise until, at about five billion degrees, the build-up of elements by fusion reaches a dead end. At this stage the build-up would form the most stable of all the elements, namely iron and its neighbors (around atomic weight 56). Any nuclear reaction involving the iron group must absorb energy rather than release it; hence these nuclei cannot serve as fuel to continue the chain of fusions.

Hoyle has suggested that this impasse may account for the anomalous abundance of the iron group of elements in the universe. As the primeval stars grow older, they accumulate iron as the end product. If they reach the stage where they have burned up all their internal fuel and then explode (perhaps as a result of a sudden disturbance of the hot core material and its reaction with unburned material in the envelope of the star), they will fling a considerable amount of iron into interstellar space.

We must now pause to relate the element-building processes to the evolution of stars. Clearly in the early stages of a star's evolution the only, or at least dominant, process is the build-up of hydrogen to helium. The fusion of hydrogen to helium is, in fact, the main source of energy of most stars (which fall in what is called the "main sequence" on the familiar chart of star classifications). Recall that most of the matter in the universe is hydrogen and helium: we can assign the building of all the other elements to comparatively minor or rare processes in the life of stars.

It is in the old "red giants" that the fusion of helium into carbon and successively heavier elements takes over the dominant role. But, as we have just seen, we have reached an impasse at iron, and

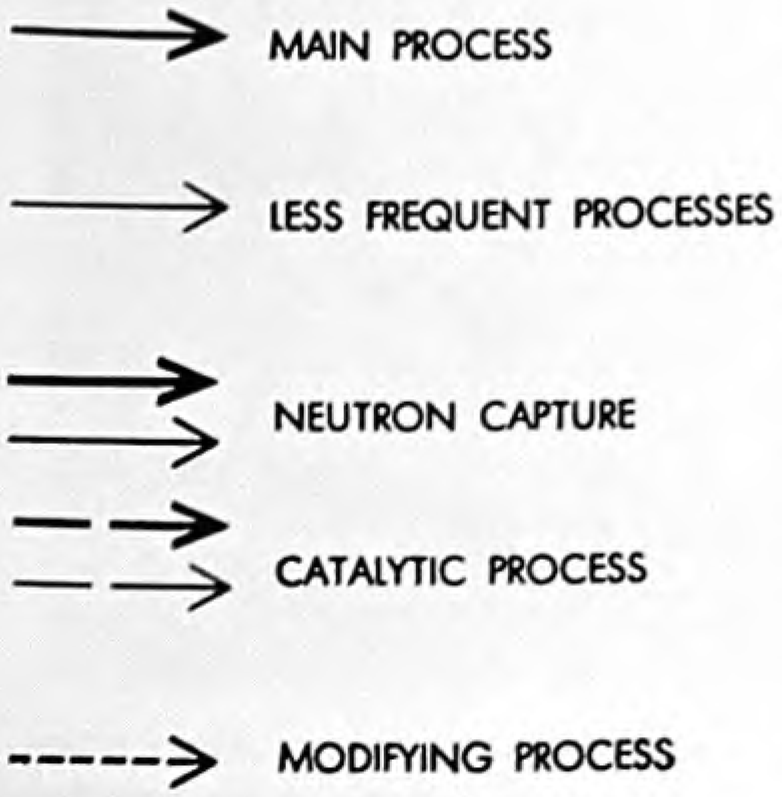


HYDROGEN BURNING
HELIUM BURNING
H
PP
CN
He 4
C 12
C 13
N 14
N 15
C 12
n
O 16
O 17
O 18
F 19
n
Ne 20
Ne 21
Ne 22
Na 23
Ne 20
n
Mg 24
S 32
NUCLEAR REACTIONS AT EQUILIBRIUM
IRON GROUP
Ti 50
HEAVY ELEMENTS
TRANSURANIUM ELEMENTS

we must now find some way to construct the elements beyond the iron group. Here Gamow's concept of build-up by neutron capture, and what we know of certain cataclysmic events in the history of stars, comes to our aid.

Stars, like human beings, are subject to accidents and disorders: not all of them live to a ripe old age. They occasionally boil up to a state of instability that results in their exploding as a nova or supernova. This may happen in a star of any age, young or old. When a young star explodes, it discharges hydrogen and helium into interstellar space. An old star will spew forth not only these nuclei but also other elements from carbon up to iron. Besides this, even stable stars, including our sun, are known to be constantly ejecting corpuscles of matter into space.

ENTIRE SCHEME of the synthesis of elements in stars is presented on the opposite page. Elements synthesized by interactions with protons (hydrogen burning) are listed horizontally. Elements synthesized by interactions with alpha particles (helium burning) and more complicated processes are listed vertically. The letters "pp" stand for the proton-proton reaction; the letters "CN," for the carbon-nitrogen cycle. The letter "n" stands for neutrons liberated in nuclear reactions and thus available for neutron-capture processes. The production of carbon (C), nitrogen (N), oxygen (O), fluorine (F), neon (Ne) and sodium (Na) are given in detail. Described in less detail are the neutron-capture processes responsible for magnesium (Mg), sulfur (S), titanium (Ti), the iron group, the heavy elements and the transuranium elements. The colored dotted line between the transuranium elements and the heavy elements represents alpha decay or fission.



Thus a debris of matter from living and dying stars pours into space, and its elements mix with the interstellar gas. From this material new stars are born: astronomy today has strong evidence of the existence of young or infant stars in the heavens. So we can postulate two kinds of stars: primeval or "first generation" stars, and "second generation" stars, which start with a legacy of the elements up to iron from the parents of their matter.

Let us now consider a second generation star which has condensed from hydrogen mixed with some carbon, oxygen, neon and even a little iron. In these stars hydrogen in the core will again be converted to helium, but now, because carbon is present, the conversion will take the route of the second process described by Bethe, the carbon-nitrogen cycle. In this cycle carbon 12 captures hydrogen nuclei in a series of steps which converts it successively to carbon 13, nitrogen 14 and nitrogen 15: in the end nitrogen 15 takes on another proton, breaks down to carbon 12 again, and in so doing emits a nucleus of helium. Thus the chain of reactions produces helium and all the isotopes of carbon and nitrogen. It can be calculated that this process, rather than the direct fusion of protons, is the source of energy in second-generation main-sequence stars which are large enough to have internal temperatures over 15 million degrees.

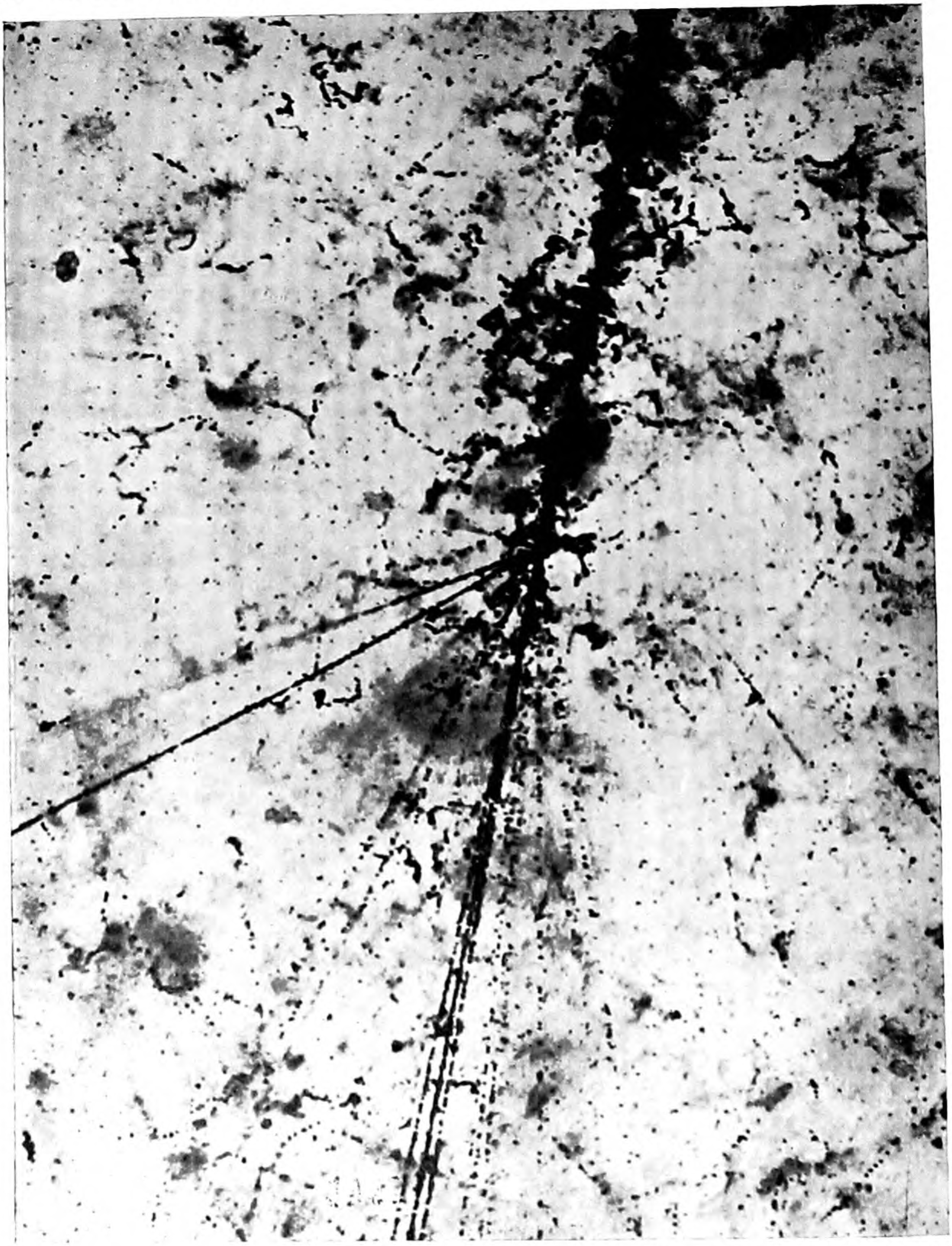
The oxygen in the star's core mixture is converted by proton capture to the isotope oxygen 17, and neon similarly to neon 21. Now these isotopes, and carbon 13, come to play a crucial role when the star arrives at the red giant stage and its core consists mainly of hot helium. The three isotopes, on reacting with helium, produce unstable nuclei which emit neutrons; so laboratory experiments have shown. Consequently they furnish a steady supply of neutrons within the core. We have seen that all nuclei, even iron, readily capture neutrons. Here, then, is the mechanism that breaks the iron bottleneck. By successive captures of neutrons, nuclei can be built up from the iron group to elements as heavy as lead and bismuth. The slow neutron-capture process in the core of a star cannot carry the build-up beyond bismuth, because the heavier elements decay too rapidly (by emitting alpha particles, or helium nuclei). However, the heavy elements *can* capture neutrons at a sufficiently rapid rate to continue the chain during an explosion of a star.

That the stars do in fact synthesize heavy elements has been confirmed by considerable evidence, some of it spectacular. The most dramatic was the discovery of the element technetium in certain giant stars (through its spectral fingerprints). Technetium is an unstable element whose longest-lived known isotope has a half-life of only 216,000 years—far less than the age of the stars in which it is found. It must therefore have been made in the star long after the star's birth. As for the synthesis of the heaviest elements, an isotope of the element californium was found in the debris from a thermonuclear explosion in the Bikini tests of 1952, and we have seen an intriguing suggestion of its presence in certain supernovae. After their original flare-up, these exploding stars decline in brightness at a rate which corresponds to a half-life (decline to half the intensity) of 55 days, and this is just the half-life of spontaneous fission of californium 254!

Research in our own and other laboratories has now established possible processes for synthesis of all the elements [*see opposite page*]. Of course this scheme is still highly tentative. It is disconcerting that so many different processes have to be invoked; it would be much more satisfying to see a single process that could build all the elements. The picture may, however, become simpler as more research is done. What particularly gratifies workers in this field is that speculation about the origin of the elements has been reduced to questions specific enough to be tested both by nuclear physicists in the laboratory and by astrophysicists studying the stars.

There is food for philosophical thought in what has been learned so far. The heavy elements, of which our solar system has its full share, took a long time to produce—probably one to two billion years. Thus the particular part of the universe we inhabit is not the oldest thing in it; many cosmic events preceded the formation of the earth. The oldest stars in our galaxy are estimated to have an age of 6.5 billion years, while analyses of meteorites indicate that the solar system is no more than 4.5 billion years old.

Copernicus displaced the center of the universe from the earth to the sun; later cosmologists dethroned the solar system as the center; now we see that our system was not even in existence at the beginning of the galaxy. So dies the last vestige of mankind's geocentric conception of the universe.



COSMIC RAY smashes a nucleus in a photographic emulsion. The character of the original track indicates that it was made by a nucleus of iron. The distribution of elements in cosmic rays provides evidence on the relative abundance of elements in the universe.

The Author

WILLIAM A. FOWLER is professor of physics at the California Institute of Technology. Before he got his Ph.D. at Cal Tech in 1936, he had been one of the first men to receive a bachelor's degree in engineering physics at Ohio State University. Ever since he came to Cal Tech in 1933, Fowler has been associated with C. C. Lauritsen, first as a student and later as a physicist in the W. K. Kellogg Radiation Laboratory, which Lauritsen directs. "During World War II we worked on rocket ordnance in Kellogg, and one of the faculty members associated with us was the astronomer I. S. Bowen, who directed all photographic measurements in our field-testing program at Goldstone Dry Lake in the Mojave Desert and later at Inyokern. After the war Dr. Bowen became director of the Mount Wilson and Palomar Observatories, and early in 1946 he held a series of informal seminars in his home on nuclear problems in astrophysics and astronomy. Lauritsen and I and our students attended, and as a result of the interest that developed we decided to study experimentally in the laboratory those particular nuclear reactions which were thought to take place in stars. Research along these lines has been a part of our laboratory program since that time. In 1948 Jesse L. Greenstein came to Cal Tech, and his interest in the abundances of the elements in stars has stimulated much of our work. Fred Hoyle and E. E. Salpeter have been frequent visiting lecturers at Cal Tech, and from them we have learned about the basic mechanisms by which nuclear processes supply energy and synthesize elements in stars." Fowler spent the 1954-1955 academic year as a Fulbright lecturer and Guggenheim fellow in the Cavendish Laboratory at the University of Cambridge.

Bibliography

ELEMENT-BUILDING REACTIONS IN STARS. William A. Fowler and Jesse L. Greenstein in *Proceedings of the National Academy of Sciences*, Vol. 42, No. 4, pages 173-180; April, 1956.

STELLAR EVOLUTION AND THE SYNTHESIS OF THE ELEMENTS. W. A. Fowler, G. R. Burbidge and E. Margaret Burbidge in *The Astrophysical Journal*, Vol. 122, No. 2, pages 271-285; September, 1955.

THEORY OF THE ORIGIN AND RELATIVE ABUNDANCE DISTRIBUTION OF THE ELEMENTS. Ralph A. Alpher and Robert C. Herman in *Reviews of Modern Physics*, Vol. 22, No. 2, pages 153-212; April, 1950.

SPECIAL NOTE TO TEACHERS: Each article in this volume, plus more than 660 others, is available as a separate, self-bound SCIENTIFIC AMERICAN Offprint. Offprints may be ordered in any combination and in any quantity. Teachers who want to adopt articles for their courses, therefore, can ensure that each student has his own set. Students' sets are collated by the publisher before shipment.

SCIENTIFIC AMERICAN September 1956, Vol. 195, No. 3, pp. 136-154 OFFPRINT 211

THE EVOLUTIONARY UNIVERSE

by George Gamow

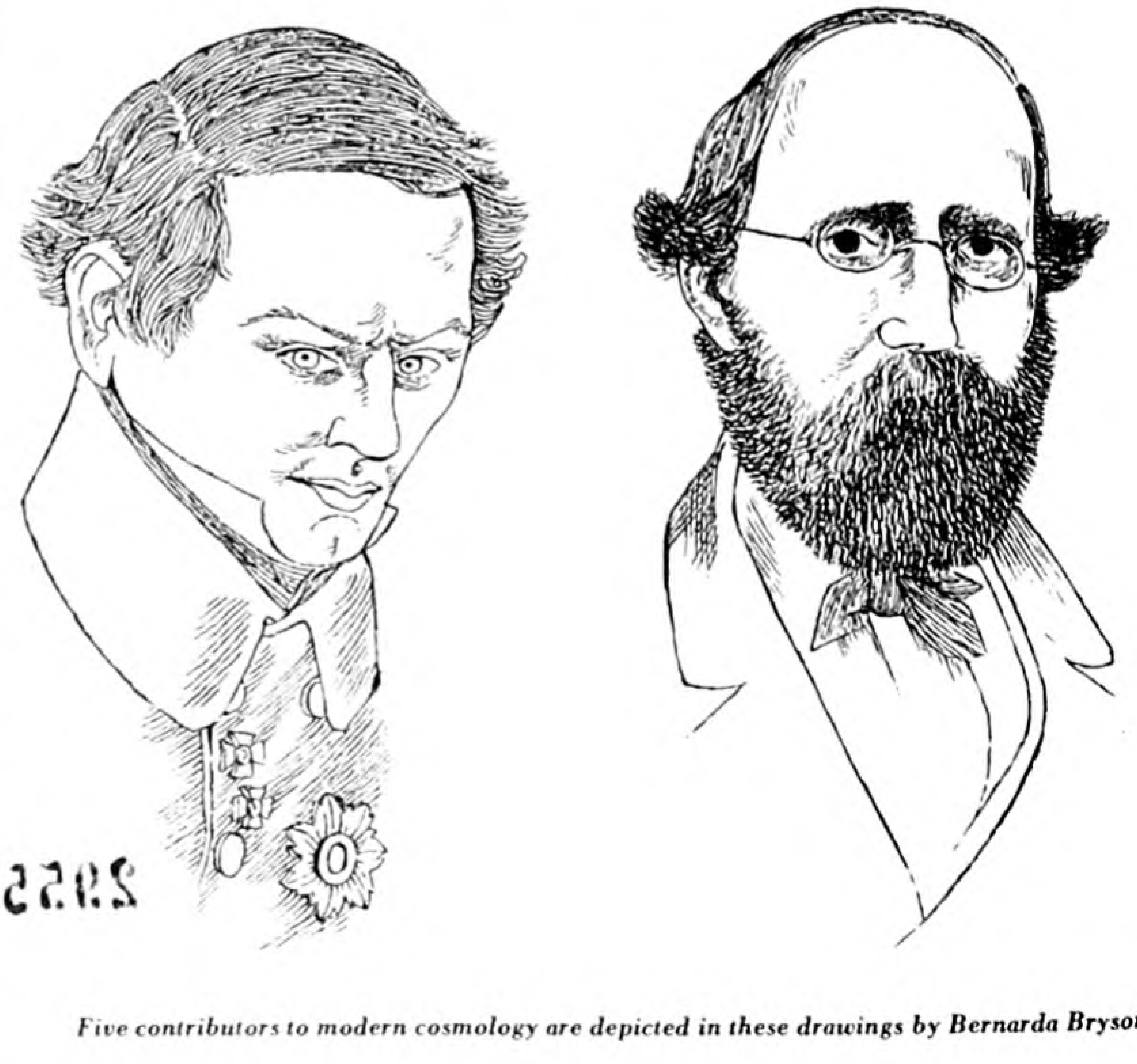
Most cosmologists believe that the universe began as a dense kernel of matter and radiant energy which started to expand about five billion years ago and later coalesced into galaxies.

Cosmology is the study of the general nature of the universe in space and in time—what it is now, what it was in the past and what it is likely to be in the future. Since the only forces at work between the galaxies that make up the material universe are the forces of gravity, the cosmological problem is closely connected with the theory of gravitation, in particular with its modern version as comprised in Albert Einstein's general theory of relativity. In the frame of this theory the properties of space, time and gravitation are merged into one harmonious and elegant picture.

The basic cosmological notion of general relativity grew out of the work of great mathematicians of the 19th century. In the middle of the last century two inquisitive mathematical minds—a Russian named Nikolai Lobachevski and a Hungarian named János Bolyai—discovered that the classical geometry of Euclid was not the only possible geometry: in fact, they succeeded in constructing a geometry which was fully as logical and self-consistent as the Euclidean. They began by overthrowing Euclid's axiom about parallel lines: namely, that only one parallel to a given straight line can be drawn through a point not on that line. Lobachevski and Bolyai both conceived a system of geometry in which a great number of lines parallel to a given line could be drawn through a point outside the line.

To illustrate the differences between Euclidean geometry and their non-Euclidean system it is simplest to consider just two dimensions—that is, the geometry of surfaces. In our schoolbooks this is known as "plane geometry," because the Euclidean surface is a flat surface. Suppose, now, we examine the properties of a two-dimensional geometry constructed not on a plane surface but on a curved surface. For the system of Lobachevski and Bolyai we must take the curvature of the surface to be "negative," which means that the curvature is not like that of the surface of a sphere but like that of a saddle [*see illustrations on page 80*]. Now if we are to draw parallel lines or any figure (*e.g.*, a triangle) on this surface, we must decide first of all how we shall define a "straight line," equivalent to the straight line of plane geometry. The most reasonable definition of a straight line in Euclidean geometry is that it is the path of the shortest distance between two points. On a curved surface the line, so defined, becomes a curved line known as a "geodesic" [see "The Straight Line," by Morris Kline; SCIENTIFIC AMERICAN, March].

Considering a surface curved like a saddle, we find that, given a "straight" line or geodesic, we can draw through a point outside that line a great many geodesics which will never intersect the given line, no matter how far they are extended. They are therefore parallel to it, by the definition of parallel. The possible parallels to the line fall within certain limits, indicated by the intersecting



Five contributors to modern cosmology are depicted in these drawings by Bernarda Bryson.

lines in the drawing at the left in the middle of the next page.

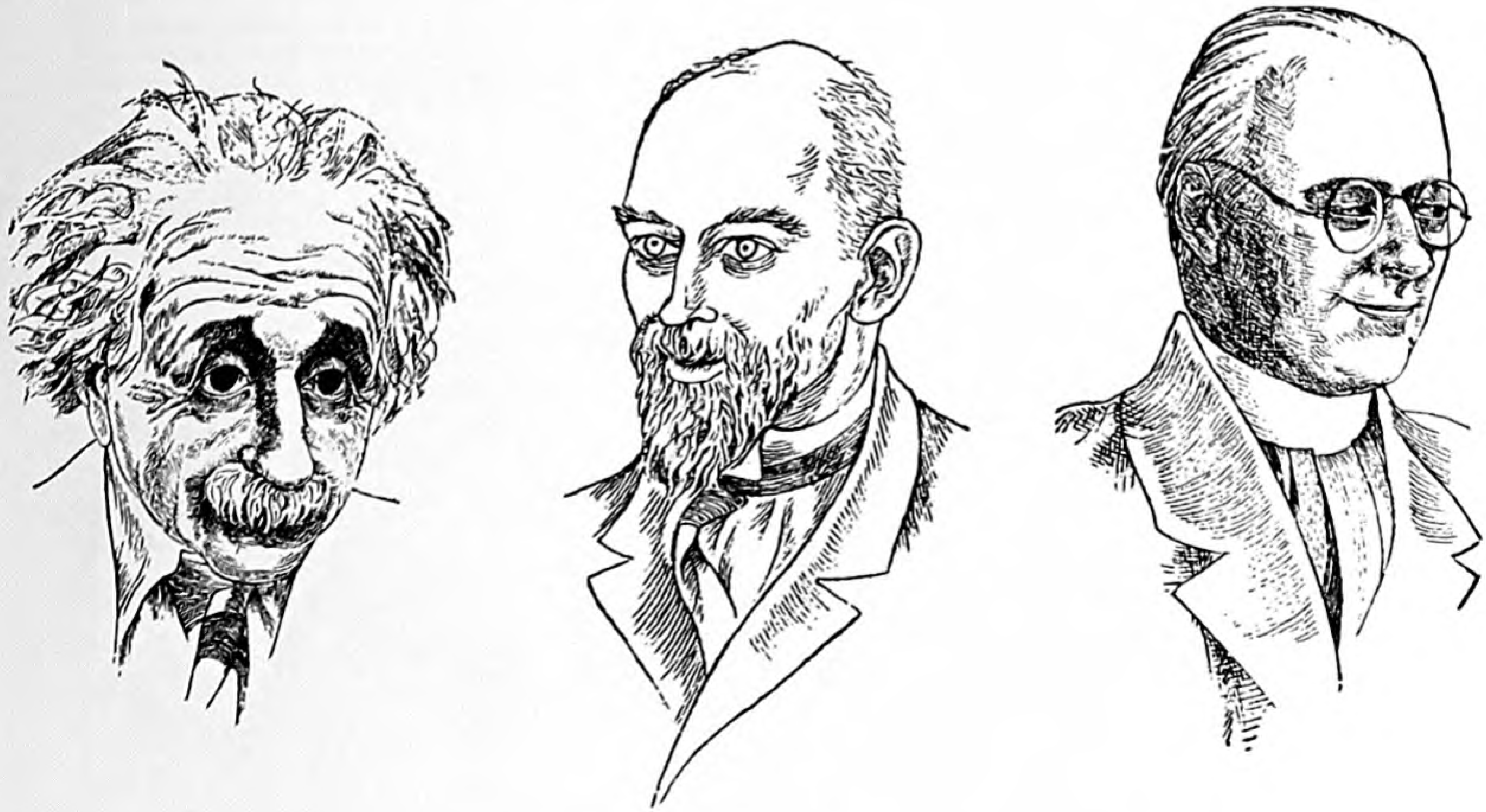
As a consequence of the overthrow of Euclid's axiom on parallel lines, many of his theorems are demolished in the new geometry. For example, the Euclidean theorem that the sum of the three angles of a triangle is 180 degrees no longer holds on a curved surface. On the saddle-shaped surface the angles of a triangle formed by three geodesics always add up to less than 180 degrees, the actual sum depending on the size of the triangle. Further, a circle on the saddle surface does not have the same properties as a circle in plane geometry. On a flat surface the circumference of a circle increases in proportion to the increase in diameter, and the area of a circle increases in proportion to the square of the increase in diameter. But on a saddle surface both the circumference and the area of a circle increase at *faster* rates than on a flat surface with increasing diameter.

After Lobachevski and Bolyai, the German mathematician Bernhard Riemann constructed another non-Euclidean geometry whose two-dimensional model is a surface of positive, rather than negative, curvature—that is, the surface of a sphere. In this case a geodesic line is simply a great circle around the sphere or a segment of such a circle, and since any two great circles must intersect at two points (the poles), there are no parallel lines at all in this geometry. Again the sum of the three angles of a triangle is not 180 degrees: in this case it is always *more* than 180. The circumference of a circle now increases at a rate *slower* than in proportion to its increase in diameter, and its area increases more slowly than the square of the diameter.

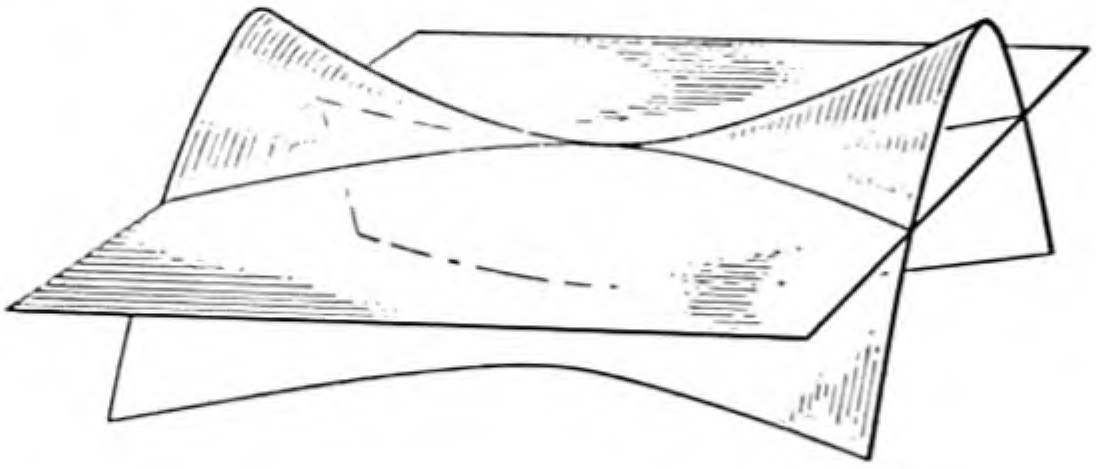
Now all this is not merely an exercise in abstract reasoning but bears directly on the geometry of the universe in which we live. Is the space of our universe "flat," as Euclid assumed, or is it curved negatively (per Lobachevski and Bolyai) or curved positively (Riemann)? If we were two-dimensional creatures living in a two-dimensional universe, we could tell whether we were living on a flat or a curved surface by studying the properties of triangles and circles drawn on that surface. Similarly as three-dimensional beings living in three-dimensional space we should be able, by studying geometrical properties of that space, to decide what the curvature of our space is. Riemann in fact developed mathematical formulas describing the properties of various kinds of curved space in three and more dimensions. In the early years of this century Einstein conceived the idea of the universe as a curved system in four dimensions, embodying time as the fourth dimension, and he proceeded to apply Riemann's formulas to test his idea.

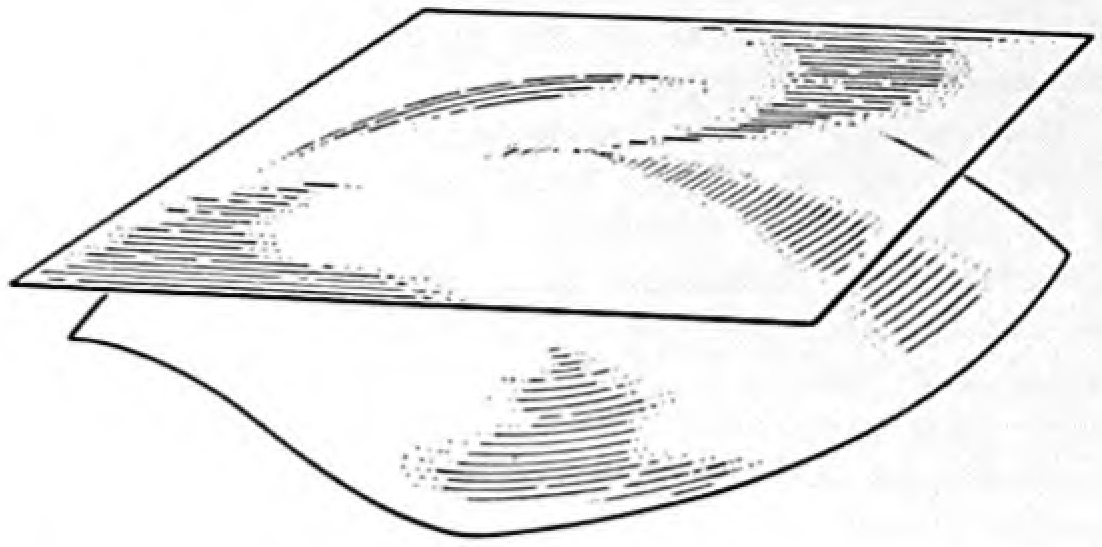
Einstein showed that time can be considered a fourth coordinate supplementing the three coordinates of space. He connected space and time, thus establishing a "space-time continuum," by means of the speed of light as a link between time and space dimensions. However, recognizing that space and time are physically different entities, he employed the imaginary number $\sqrt{-1}$, or i, to express the unit of time mathematically and make the time coordinate formally equivalent to the three coordinates of space.

In his special theory of relativity Einstein made the geometry of the time-space continuum strictly Euclidean, that is, flat. The great idea that he introduced later in his general theory was that gravitation, whose effects had been neglected in the special theory, must make it curved. He saw that the gravitational effect of the masses distributed in space and moving in time was equivalent to curvature of the four-dimensional space-time continuum. In place of the classical Newtonian statement that "the sun produces a field of forces which impels the earth to deviate from straight-line mo-

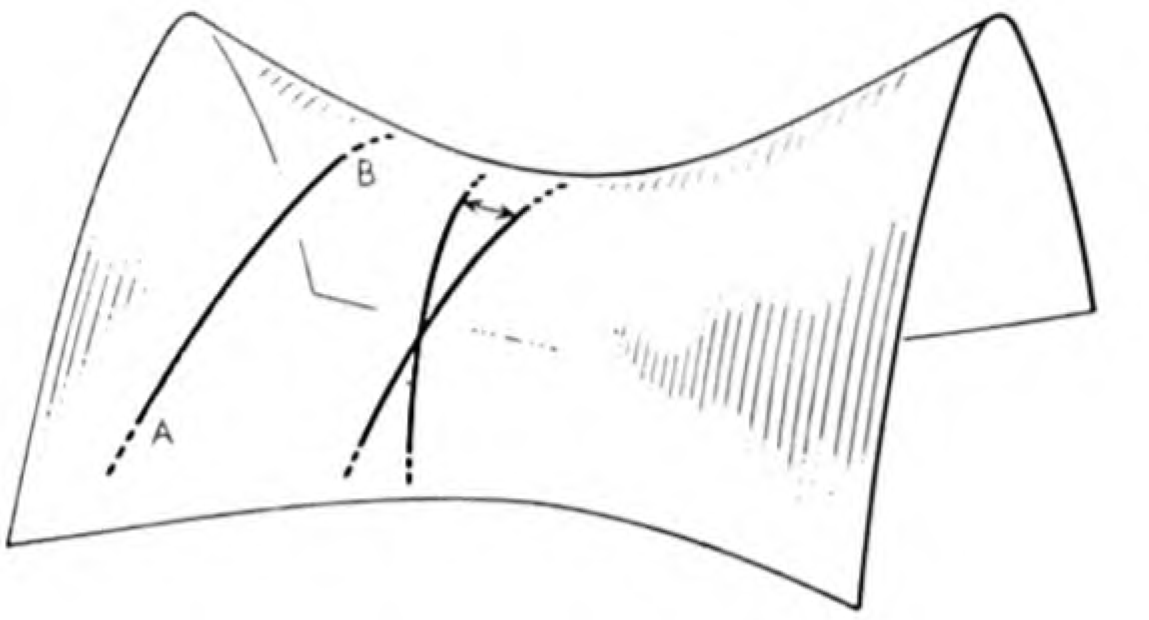


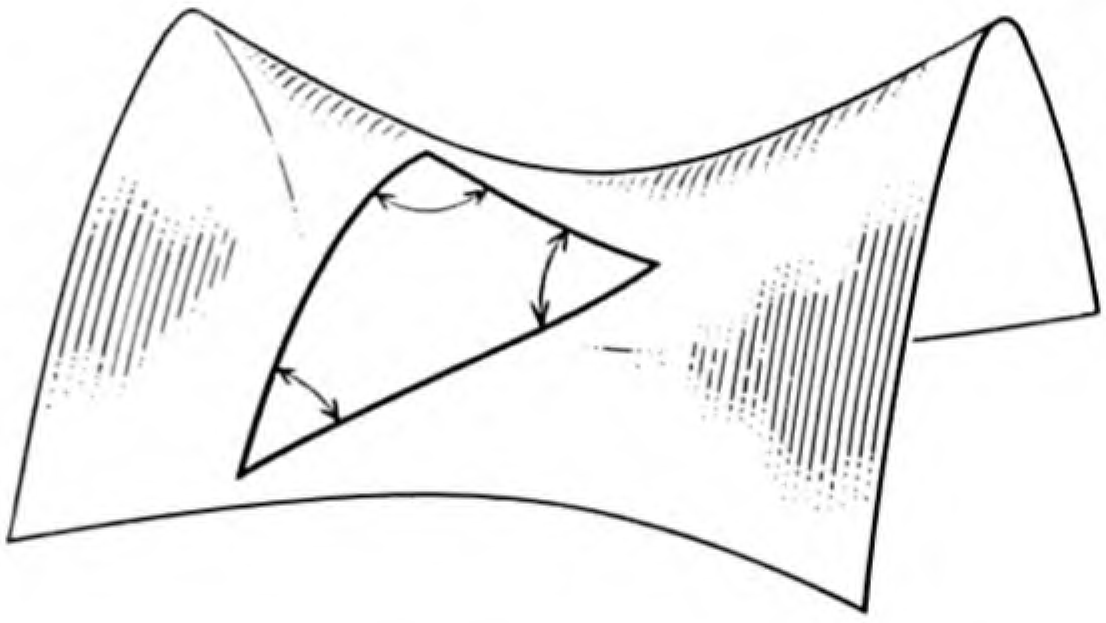
From left to right they are: Nikolai Lobachevski, Bernhard Riemann, Albert Einstein, Willem de Sitter and Georges Lemaître



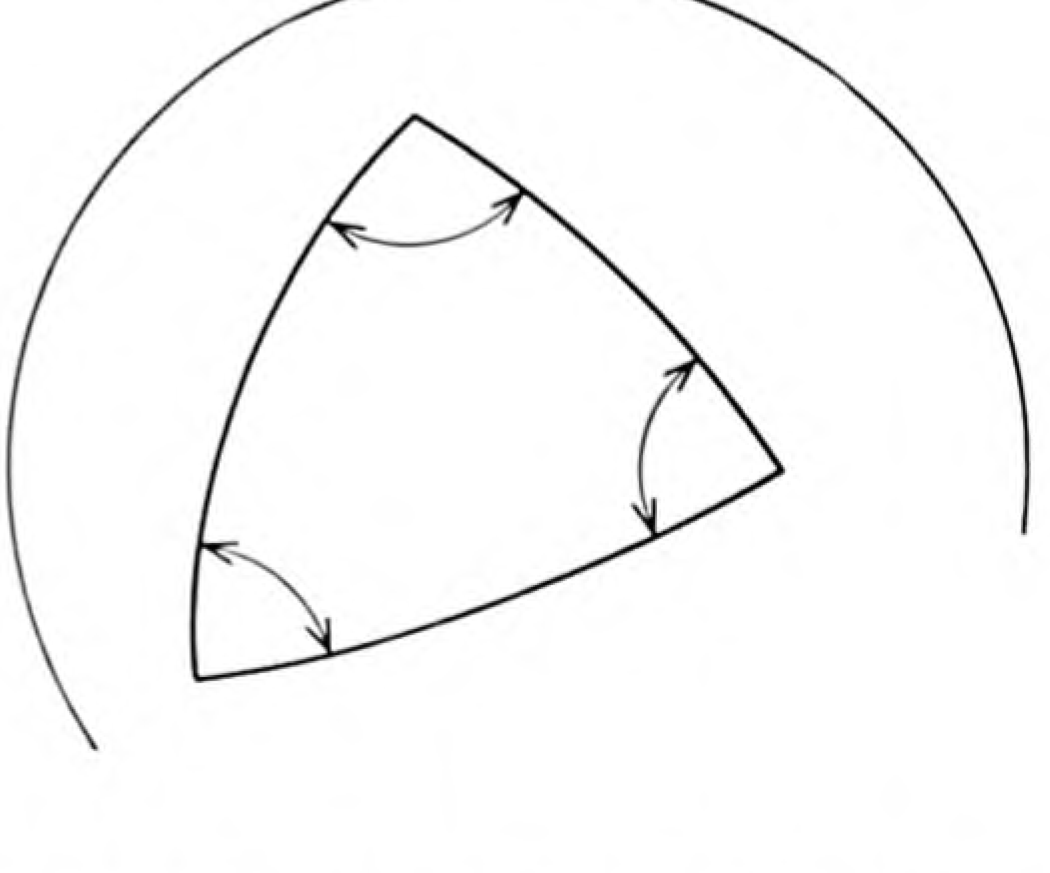


NEGATIVE AND POSITIVE CURVATURE of space is suggested by this two-dimensional analogy. The saddle-shaped surface at left, which lies on both sides of a tangential plane, is negatively curved. The spherical surface at right, which lies on one side of a tangential plane, is positively curved. If space is negatively curved, the universe is infinite; if it is positively curved, the universe is finite.





ON A NEGATIVELY CURVED SURFACE the shortest distance between two points is not a straight line but a curved "geodesic," such as the line AB at the left. On a plane surface only one parallel to a given straight line can be drawn through a point not on that line; on a negatively curved surface many geodesics can be drawn through a point not on a given geodesic without ever intersecting it. These "parallel" lines will fall within the limits indicated by the arrow between the intersecting lines at left. On a plane surface the angles of a triangle add up to 180 degrees; on the negatively curved surface at the right, they add up to less than 180 degrees.



ON A POSITIVELY CURVED SURFACE the shortest distance between two points follows a great circle, a closed line passing through opposite points on the surface (*single curved line at left*). In this geometry there are no "parallel" lines because any two great circles must intersect. The circumference of a circle increases more slowly with diameter than on a flat surface, and the area similarly increases more slowly (*concentric circles at left*). The angles of a triangle on the surface (*right*) add up to more than 180 degrees.

tion and to move in a circle around the sun." Einstein substituted a statement to the effect that "the presence of the sun causes a curvature of the space-time continuum in its neighborhood."

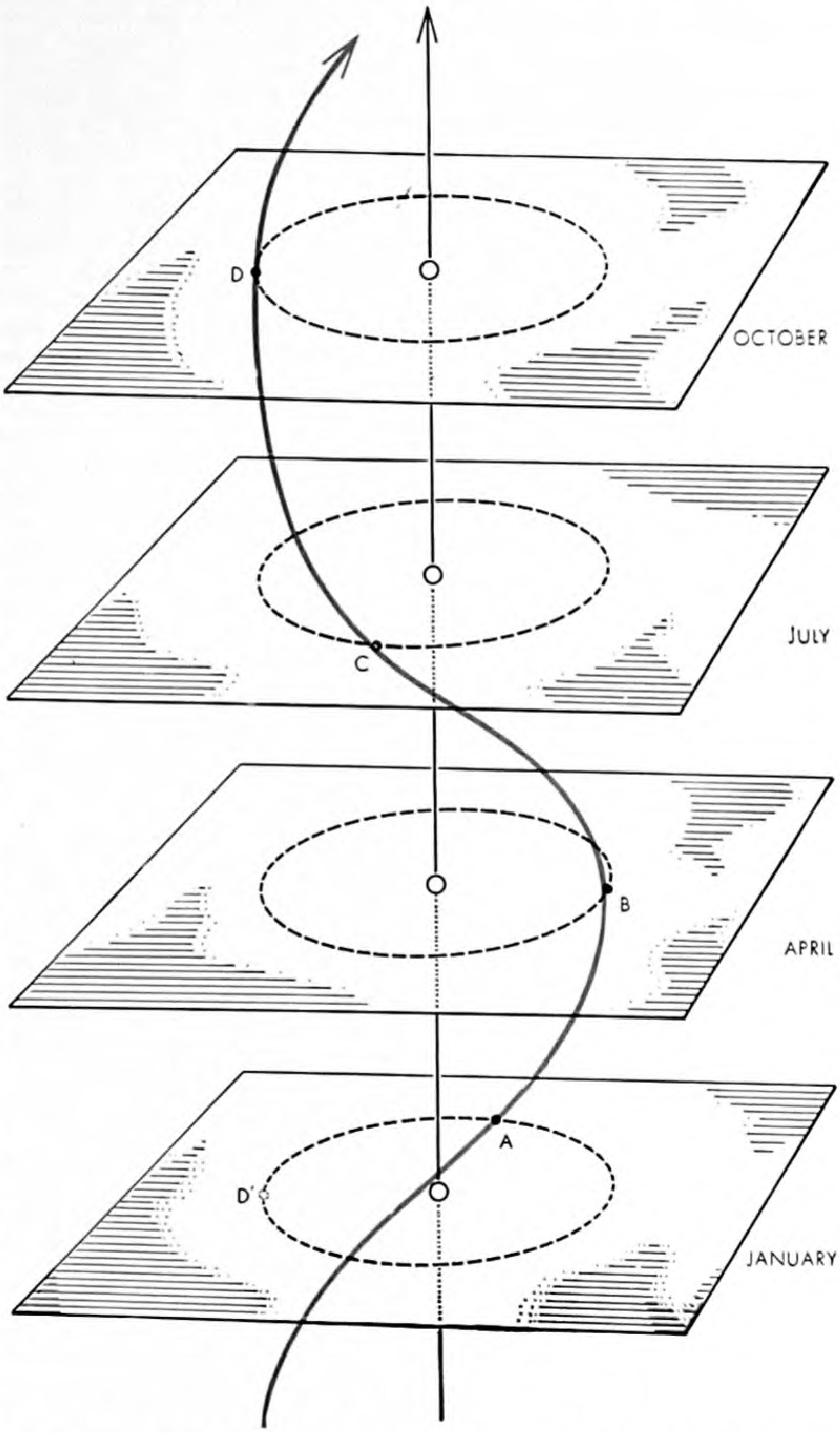
The motion of an object in the space-time continuum can be represented by a curve called the object's "world line." For example, the world line of the earth's travel around the sun in time is pictured in the drawing on the next page. (Space must be represented here in only two dimensions; it would be impossible for a three-dimensional artist to draw the fourth dimension in this scheme, but since the orbit of the earth around the sun lies in a single plane, the omission is unimportant.) Einstein declared, in effect: "The world line of the earth is a geodesic in the curved four-dimensional space around the sun." In other words, the line ABCD in the drawing corresponds to the shortest *four-dimensional* distance between the position of the earth in January (at A) and its position in October (at D).

Einstein's idea of the gravitational curvature of space-time was, of course, triumphantly affirmed by the discovery of perturbations in the motion of Mercury at its closest approach to the sun and of the deflection of light rays by the sun's gravitational field. Einstein next attempted to apply the idea to the universe as a whole. Does it have a general curvature, similar to the local curvature in the sun's gravitational field? He now had to consider not a single center of gravitational force but countless centers of attraction in a universe full of matter concentrated in galaxies whose distribution fluctuates considerably from region to region in space. However, in the large-scale view the galaxies are spread fairly uniformly throughout space as far out as our biggest telescopes can see, and we can justifiably "smooth out" its matter to a general average (which comes to about one hydrogen atom per cubic meter). On this assumption the universe as a whole has a smooth general curvature.

But if the space of the universe is curved, what is the sign of this curvature? Is it positive, as in our two-dimensional analogy of the surface of a sphere, or is it negative, as in the case of a saddle surface? And, since we cannot consider space alone, how is this space curvature related to time?

Analyzing the pertinent mathematical equations, Einstein came to the conclusion that the curvature of space must be independent of time, *i.e.*, that the universe as a whole must be unchanging (though it changes internally). However, he found to his surprise that there was no solution of the equations that would permit a static cosmos. To repair the situation, Einstein was forced to introduce an additional hypothesis which amounted to the assumption that a new kind of force was acting among the galaxies. This hypothetical force had to be independent of mass (being the same for an apple, the moon and the sun!) and to gain in strength with increasing distance between the interacting objects (as no other forces ever do in physics!).

Einstein's new force, called "cosmic repulsion," allowed two mathematical models of a static universe. One solution, which was worked out by Einstein him-



MOTION OF BODY in the curved "space-time continuum" of Albert Einstein is represented by the "world line" of the earth's motion around the sun. Here the sun is the small open circle in each of the four planes. The earth is the black dot on the elliptical orbit. Each plane shows the position of the earth at a month of the year. The world line is in color.

self and became known as "Einstein's spherical universe," gave the space of the cosmos a positive curvature. Like a sphere, this universe was closed and thus had a finite volume. The space coordinates in Einstein's spherical universe were curved in the same way as the latitude or longitude coordinates on the surface of the earth. However, the time axis of the space-time continuum ran quite straight, as in the good old classical physics. This means that no cosmic event would ever recur. The two-dimensional analogy of Einstein's space-time continuum is the surface of a cylinder, with the time axis running parallel to the axis of the cylinder and the space axis perpendicular to it [*see drawing at left on this page*].

The other static solution based on the mysterious repulsion forces was discovered by the Dutch mathematician Willem de Sitter. In his model of the universe both space and time were curved. Its geometry was similar to that of a globe, with longitude serving as the space coordinate and latitude as time [*drawing at right on this page*].

Unhappily astronomical observations contradicted both Einstein's and de Sitter's static models of the universe, and they were soon abandoned.

In the year 1922 a major turning point came in the cosmological problem. A Russian mathematician, Alexander A. Friedman (from whom the author of this article learned his relativity), discovered an error in Einstein's proof for a static universe. In carrying out his proof Einstein had divided both sides of an equation by a quantity which, Friedman found, could become zero under certain circumstances. Since division by zero is not permitted in algebraic computations, the possibility of a nonstatic universe could not be excluded under the circumstances in question. Friedman showed that two nonstatic models were possible. One pictured the universe as expanding with time; the other, contracting.

Einstein quickly recognized the importance of this discovery. In the last edition of his book *The Meaning of Relativity* he wrote: "The mathematician Friedman found a way out of this dilemma. He showed that it is possible, according to the field equations, to have a finite density in the whole (three-dimensional) space, without enlarging these field equations ad hoc." Einstein remarked to me many years ago that the cosmic repulsion idea was the biggest blunder he had made in his entire life.

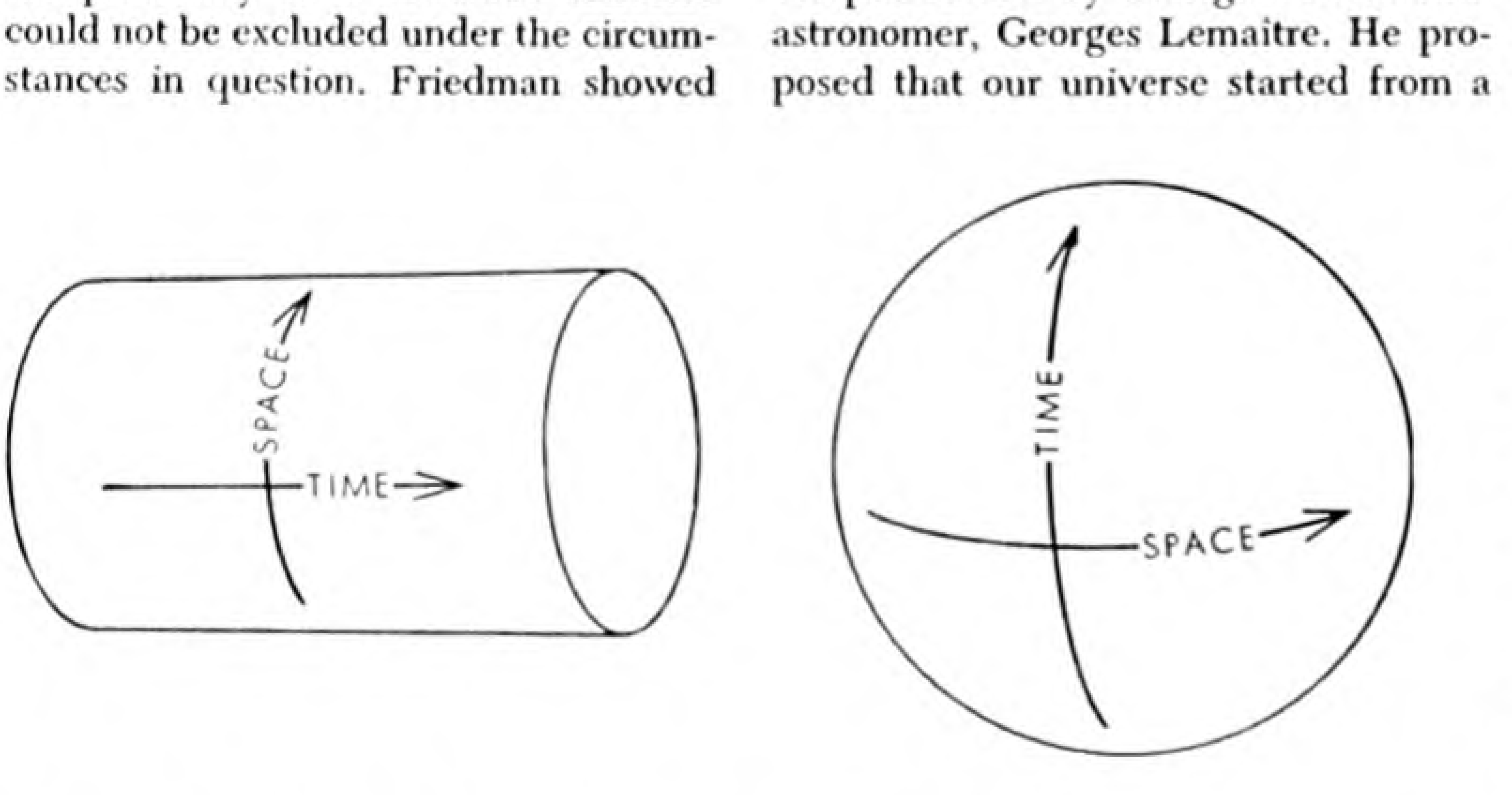
Almost at the very moment that Friedman was discovering the possibility of an expanding universe by mathematical reasoning, Edwin P. Hubble at the Mount Wilson Observatory on the other side of the world found the first evidence of actual physical expansion through his telescope. He made a compilation of the distances of a number of far galaxies, whose light was shifted toward the red end of the spectrum, and it was soon found that the extent of the shift was in direct proportion to a galaxy's distance from us, as estimated by its faintness. Hubble and others interpreted the red-shift as the Doppler effect—the well-known phenomenon of lengthening of wavelengths from any radiating source that is moving rapidly away (a train whistle, a source of light or whatever). To date there has been no other reasonable explanation of the galaxies' red-shift. If the explanation is correct, it means that the galaxies are all moving away from one another with increasing velocity as they move farther apart.

Thus Friedman and Hubble laid the foundation for the theory of the expanding universe. The theory was soon developed further by a Belgian theoretical astronomer, Georges Lemaître. He proposed that our universe started from a highly compressed and extremely hot state which he called the "primeval atom." (Modern physicists would prefer the term "primeval nucleus.") As this matter expanded, it gradually thinned out, cooled down and reaggregated in stars and galaxies, giving rise to the highly complex structure of the universe as we know it today.

Until a few years ago the theory of the expanding universe lay under the cloud of a very serious contradiction. The measurements of the speed of flight of the galaxies and their distances from us indicated that the expansion had started about 1.8 billion years ago. On the other hand, measurements of the age of ancient rocks in the earth by the clock of radioactivity (*i.e.*, the decay of uranium to lead) showed that some of the rocks were at least three billion years old; more recent estimates based on other radioactive elements raise the age of the earth's crust to almost five billion years. Clearly a universe 1.8 billion years old could not contain five-billion-year-old rocks! Happily the contradiction has now been disposed of by Walter Baade's recent discovery that the distance yardstick (based on the periods of variable stars) was faulty and that the distances between galaxies are more than twice as great as they were thought to be. This change in distances raises the age of the universe to five billion years or more.

Friedman's solution of Einstein's cosmological equation, as I mentioned, permits two kinds of universe. We can call one the "pulsating" universe. This model says that when the universe has reached a certain maximum permissible expansion, it will begin to contract; that it will shrink until its matter has been compressed to a certain maximum density, possibly that of atomic nuclear material, which is a hundred million million times denser than water; that it will then begin to expand again—and so on through the cycle *ad infinitum*. The other model is a "hyperbolic" one: it suggests that from an infinitely thin state an eternity ago the universe contracted until it reached the maximum density, from which it rebounded to an unlimited expansion which will go on indefinitely in the future.

The question whether our universe is actually "pulsating" or "hyperbolic" should be decidable from the present rate of its expansion. The situation is analogous to the case of a rocket shot from the surface of the earth. If the velocity of the rocket is less than seven miles per second—the "escape velocity"—the rocket will climb only to a certain



SPHERICAL UNIVERSE of Einstein may be represented in two dimensions by a cylinder (*left*). Its space coordinates were positively curved but its time coordinate was straight. The spherical universe of Willem de Sitter had positively curved coordinates (*right*).

height and then fall back to the earth. (If it were completely elastic, it would bounce up again, etc., etc.) On the other hand, a rocket shot with a velocity of more than seven miles per second will escape from the earth's gravitational field and disappear in space. The case of the rèceding system of galaxies is very similar to that of an escape rocket, except that instead of just two interacting bodies (the rocket and the earth) we have an unlimited number of them escaping from one another. We find that the galaxies are fleeing from one another at seven times the velocity necessary for mutual escape.

Thus we may conclude that our universe corresponds to the "hyperbolic" model, so that its present expansion will never stop. We must make one reservation. The estimate of the necessary escape velocity is based on the assumption that practically all the mass of the universe is concentrated in galaxies. If intergalactic space contained matter whose total mass was more than seven times that in the galaxies, we would have to reverse our conclusion and decide that the universe is pulsating. There has been no indication so far, however, that any matter exists in intergalactic space, and it could have escaped detection only if it were in the form of pure hydrogen gas, without other gases or dust.

Is the universe finite or infinite? This resolves itself into the question: Is the curvature of space positive or negative—closed like that of a sphere, or open like that of a saddle? We can look for the answer by studying the geometrical properties of its three-dimensional space, just as we examined the properties of figures on two-dimensional surfaces. The most convenient property to investigate astronomically is the relation between the volume of a sphere and its radius.

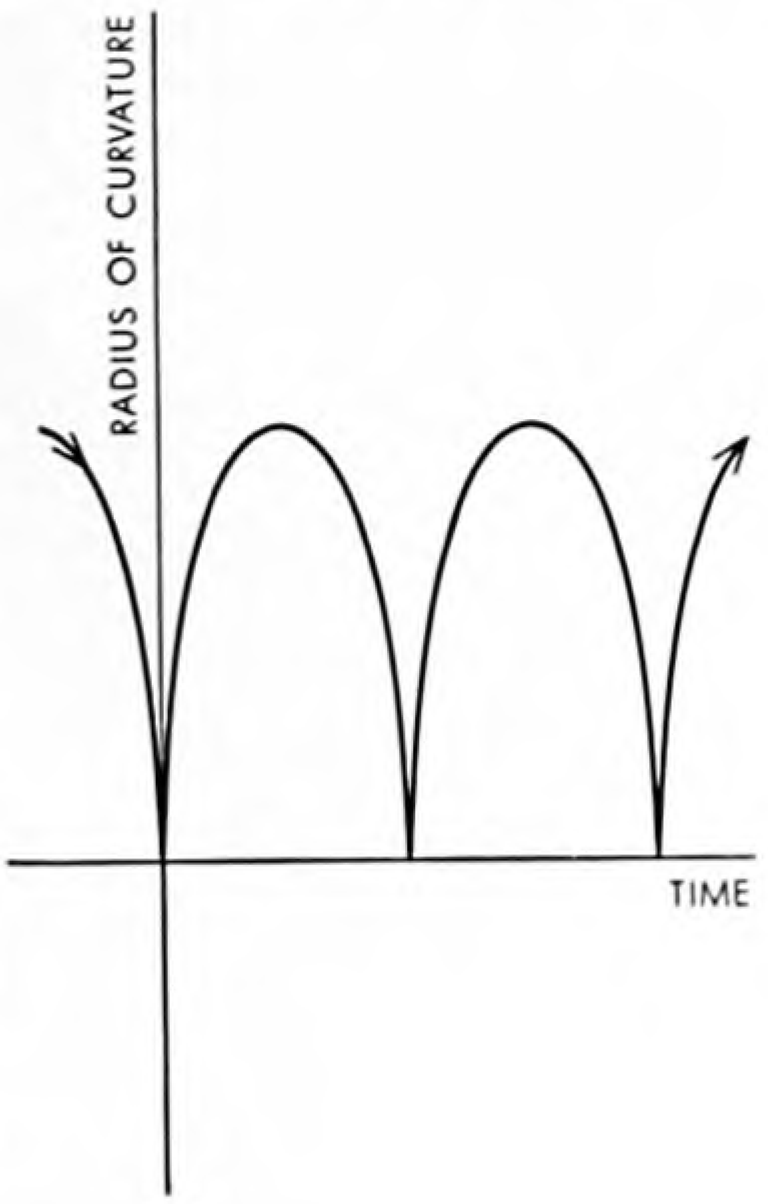
We saw that, in the two-dimensional case, the area of a circle increases with increasing radius at a faster rate on a negatively curved surface than on a Euclidean or flat surface; and that on a positively curved surface the relative rate of increase is slower. Similarly the increase of volume is faster in negatively curved space, slower in positively curved space. In Euclidean space the volume of a sphere would increase in proportion to the cube, or third power, of the increase in radius. In negatively curved space the volume would increase faster than this; in positively curved space, slower. Thus if we look into space and find that the volume of successively larger spheres, as measured by a count of the galaxies within them, increases faster than the cube of the distance to the limit of the sphere (the radius), we can conclude that the space of our universe has negative curvature, and therefore is open and infinite. By the same token, if the number of galaxies increases at a rate slower than the cube of the distance, we live in a universe of positive curvature—closed and finite.

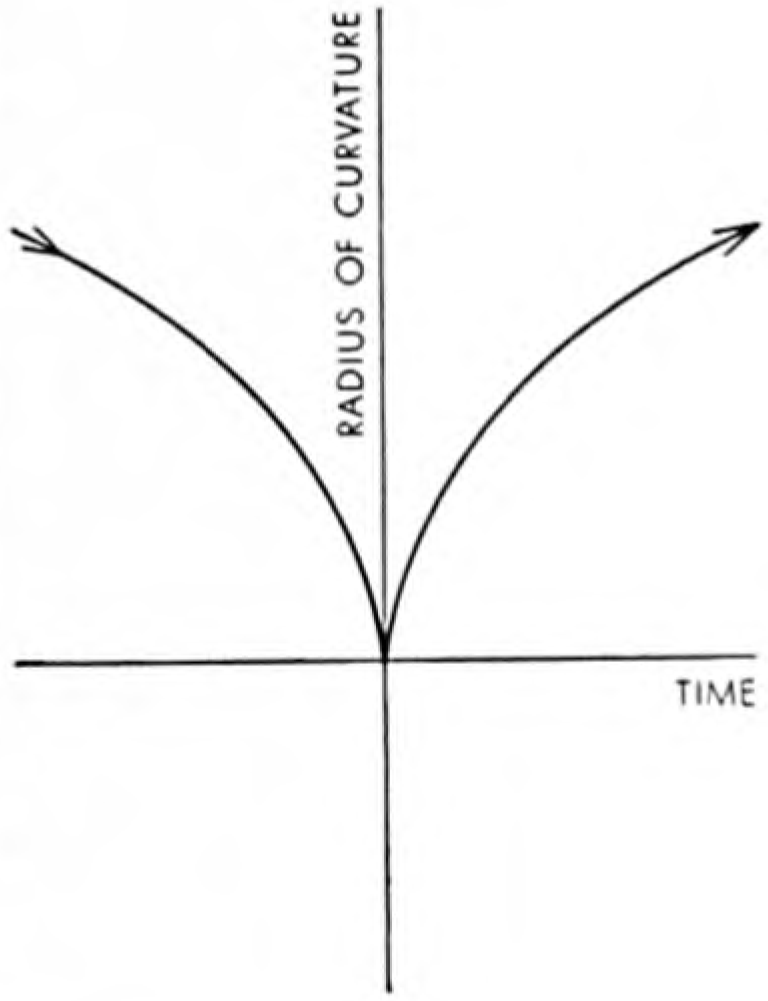
Following this idea, Hubble undertook to study the increase in number of galaxies with distance. He estimated the distances of the remote galaxies by their relative faintness: galaxies vary considerably in intrinsic brightness, but over a very large number of galaxies these variations are expected to average out. Hubble's calculations produced the conclusion that the universe is a closed system—a small universe only a few billion light-years in radius!

We know now that the scale he was using was wrong: with the new yardstick the universe would be more than twice as large as he calculated. But there is a more fundamental doubt about his result. The whole method is based on the assumption that the intrinsic brightness of a galaxy remains constant. What if it changes with time? We are seeing the light of the distant galaxies as it was emitted at widely different times in the past—500 million, a billion, two billion years ago. If the stars in the galaxies are burning out, the galaxies must dim as they grow older. A galaxy two billion light-years away cannot be put on the same distance scale with a galaxy 500 million light-years away unless we take into account the fact that we are seeing the nearer galaxy at an older, and less bright, age. The remote galaxy is farther away than a mere comparison of the luminosity of the two would suggest.

When a correction is made for the assumed decline in brightness with age, the more distant galaxies are spread out to farther distances than Hubble assumed. In fact, the calculations of volume are changed so drastically that we may have to reverse the conclusion about the curvature of space. We are not sure, because we do not yet know enough about the evolution of galaxies. But if we find that galaxies wane in intrinsic brightness by only a few per cent in a billion years, we shall have to conclude that space is curved negatively and the universe is infinite.

Actually there is another line of reasoning which supports the side of infinity. Our universe seems to be hyperbolic and ever-expanding. Mathematical solutions of fundamental cosmological equations indicate that such a universe is open and infinite.





PULSATING AND HYPERBOLIC universes are represented by curves. The pulsating universe at the top repeatedly expands to a maximum permissible density and contracts to a minimum permissible density. The hyperbolic universe at the bottom contracts and then expands indefinitely.

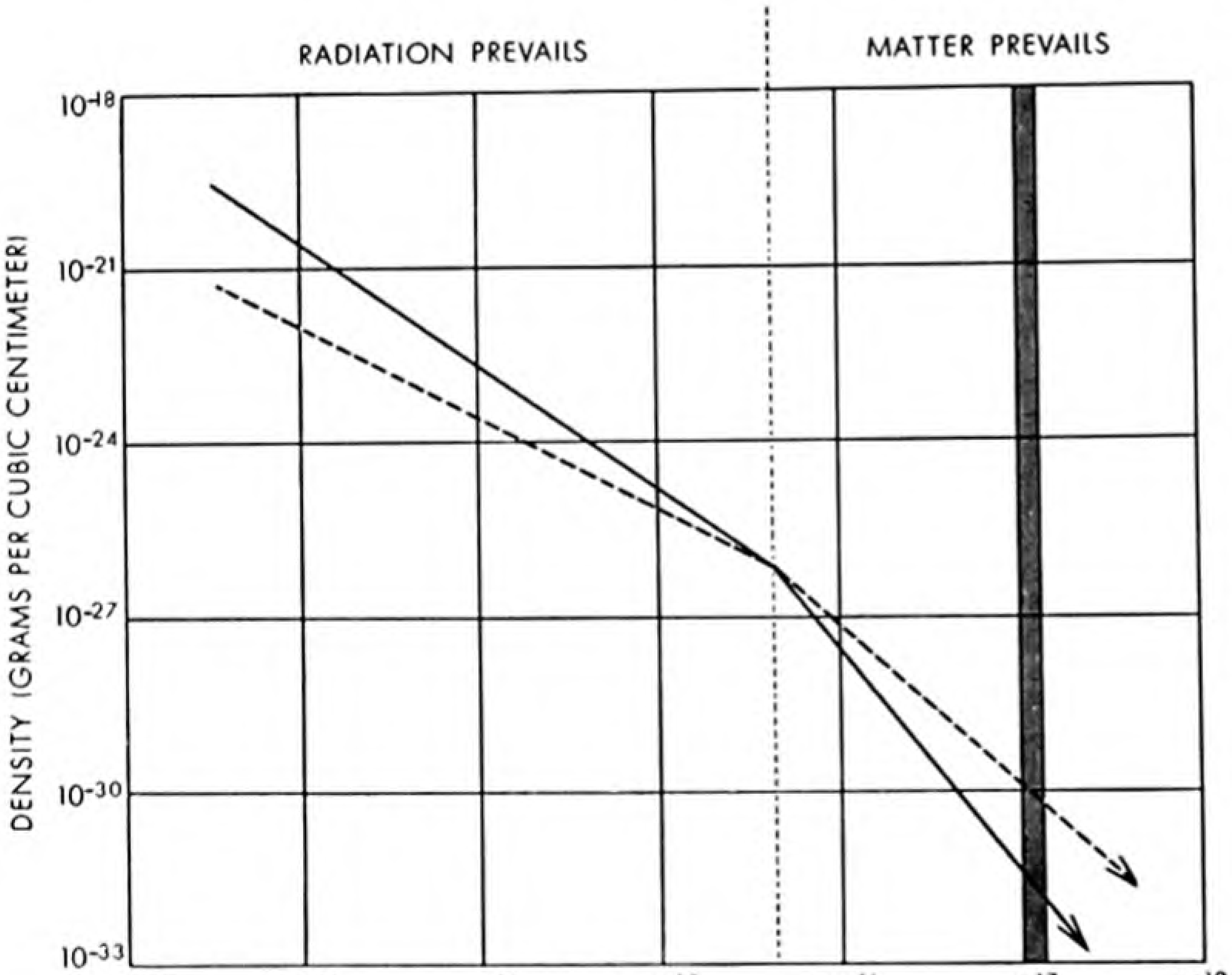
We have reviewed the questions that dominated the thinking of cosmologists during the first half of this century: the conception of a four-dimensional space-time continuum, of curved space, of an expanding universe and of a cosmos which is either finite or infinite. Now we must consider the major present issue in cosmology: Is the universe in truth evolving, or is it in a steady state of equilibri-

um which has always existed and will go on through eternity? Most cosmologists take the evolutionary view. But in 1951 a group at the University of Cambridge, whose chief spokesman has been Fred Hoyle, advanced the steady-state idea. Essentially their theory is that the universe is infinite in space and time, that it has neither a beginning nor an end, that the density of its matter remains constant, that new matter is steadily being created in space at a rate which exactly compensates for the thinning of matter by expansion, that as a consequence new galaxies are continually being born, and that the galaxies of the universe therefore range in age from mere youngsters to veterans of 5, 10, 20 and more billions of years. In my opinion this theory must be considered very questionable because of the simple fact (apart from other reasons) that the galaxies in our neighborhood all seem to be of the same age as our own Milky Way. But the issue is many-sided and fundamental, and can be settled only by extended study of the universe as far as we can observe it. Here I shall summarize the evolutionary theory.

We assume that the universe started from a very dense state of matter. In the early stages of its expansion, radiant energy was dominant over the mass of matter. We can measure energy and matter on a common scale by means of the well-known equation $E=mc^2$, which says that the energy equivalent of matter is the mass of the matter multiplied by the square of the velocity of light. Energy can be translated into mass, conversely, by dividing the energy quantity by c^2. Thus we can speak of the "mass density" of energy. Now at the beginning the mass density of the radiant energy was incomparably greater than the density of the matter in the universe. But in an expanding system the density of radiant energy decreases faster than does the density of matter. The former thins out as the fourth power of the distance of expansion: as the radius of the system doubles, the density of radiant energy drops to one sixteenth. The density of matter declines as the third power; a doubling of the radius means an eightfold increase in volume, or eightfold decrease in density.

Assuming that the universe at the beginning was under absolute rule by radiant energy, we can calculate that the temperature of the universe was 250 million degrees when it was one hour old, dropped to 6,000 degrees (the present temperature of our sun's surface) when it was 200,000 years old and had fallen to about 100 degrees below the freezing point of water when the universe reached its 250-millionth birthday.

This particular birthday was a crucial one in the life of the universe. It was the point at which the density of ordinary matter became greater than the mass density of radiant energy, because of the more rapid fall of the latter [*see chart on this page*]. The switch from the reign of radiation to the reign of matter profoundly changed matter's behavior. During the eons of its subjugation to the will of radiant energy (*i.e.*, light), it must have been spread uniformly through space in the form of thin gas. But as soon as matter became gravitationally more important than the radiant energy, it began to acquire a more interesting character. James Jeans, in his classic studies of the physics of such a situation, proved half a century ago that a gravitating gas filling a very large volume is bound to break up into individual "gas balls," the size of which is determined by the density and the temperature of the gas. Thus in the year 250,000,000 A. B. E. (after the beginning of expansion), when matter was freed from the dictatorship of radiant energy, the gas broke up into giant gas clouds, slowly drifting apart as the universe continued to expand. Applying Jeans's mathematical formula for the process to the gas filling the universe at that time, I have found that these primordial balls of gas would have had just about the mass that the galaxies of stars possess today. They were then only "protogalaxies"—cold, dark and chaotic. But their gas soon condensed into stars and formed the galaxies as we see them now.

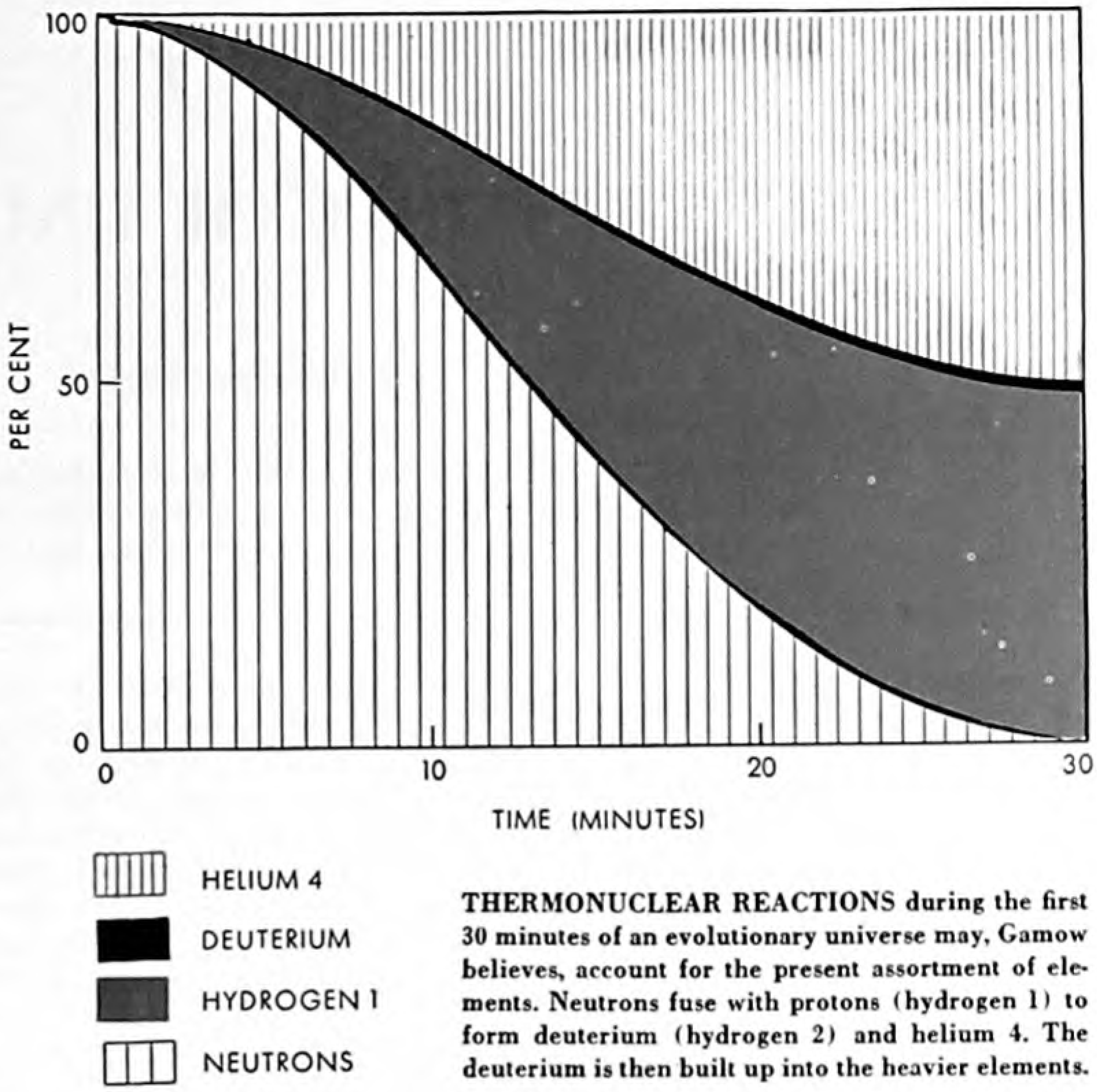


RELATIVE DENSITY OF MATTER AND RADIATION is reversed during the history of an evolutionary universe. Up to 250 million years (*broken vertical line*) the mass density of radiation (*solid curve*) is greater than that of matter (*broken curve*). After that the density of matter is greater, permitting the formation of huge gas clouds. The gray line is the present.

A central question in this picture of the evolutionary universe is the problem of accounting for the formation of the varied kinds of matter composing it—i.e., the chemical elements. My belief is that at the start matter was composed simply of protons, neutrons and electrons. After five minutes the universe must have cooled enough to permit the aggregation of protons and neutrons into larger units, from deuterons (one neutron and one proton) up to the heaviest elements. This process must have ended after about 30 minutes, for by that time the temperature of the expanding universe must have dropped below the threshold of thermonuclear reactions among light elements, and the neutrons must have been used up in element-building or been converted to protons.

To many a reader the statement that the present chemical constitution of our universe was decided in half an hour five billion years ago will sound nonsensical. But consider a spot of ground on the atomic proving ground in Nevada where an atomic bomb was exploded three years ago. Within one microsecond the nuclear reactions generated by the bomb produced a variety of fission products. Today, 100 million million microseconds later, the site is still "hot" with the surviving fission products. The ratio of one microsecond to three years is the same as the ratio of half an hour to five billion years! If we can accept a time ratio of this order in the one case, why not in the other?

The late Enrico Fermi and Anthony L. Turkevich at the Institute for Nuclear Studies of the University of Chicago undertook a detailed study of thermonuclear reactions such as must have taken place during the first half hour of the universe's expansion. They concluded that the reactions would have produced about equal amounts of hydrogen and helium, making up 99 per cent of the total material, and about 1 per cent of deuterium. We know that hydrogen and helium do in fact make up about 99 per cent of the matter of the universe. This leaves us with the problem of building the heavier elements. I hold to the opinion that some of them were built by capture of neutrons. However, since the absence of any stable nucleus of atomic weight 5 makes it improbable that the heavier elements could have been produced in the first half hour in the abundances now observed, I would agree that the lion's share of the heavy elements may well have been formed later in the hot interiors of stars.



THERMONUCLEAR REACTIONS during the first 30 minutes of an evolutionary universe may, Gamow believes, account for the present assortment of elements. Neutrons fuse with protons (hydrogen 1) to form deuterium (hydrogen 2) and helium 4. The deuterium is then built up into the heavier elements.

All the theories—of the origin, age, extent, composition and nature of the universe—are becoming more and more subject to test by new instruments and new techniques. In the course of his red-shift investigations, Allan Sandage has reported a tentative finding that the expansion of the universe may be slowing down. If this is confirmed, it may indicate that we live in a pulsating universe. But we must not forget that the estimate of distances of the galaxies is still founded on the debatable assumption that the brightness of galaxies does not change with time. If galaxies actually diminish in brightness as they age, the calculations cannot be depended upon. Thus the question whether evolution is or is not taking place in the galaxies is of crucial importance at the present stage of our outlook on the universe.

The Author

GEORGE GAMOW was born in Odessa, Russia, in 1904. As a graduate student of physics at the University of Leningrad in 1928, he found himself in hot water because "the research topic proposed by my professor was so boring that the work hardly made any progress." That summer Gamow took some courses at the University of Göttingen. While there he devised a quantum theory of radioactivity which shed new light on the structure of the atomic nucleus. Instead of returning to Leningrad he accepted an invitation from Niels Bohr, who had learned of his work, to spend a year at the University of Copenhagen on a Fellowship provided by the Carlsberg brewing firm. The following year he worked with Ernest Rutherford at the University of Cambridge on a Rockefeller fellowship. At Cambridge he wrote his first book, entitled *Constitution of Atomic Nuclei and Radioactivity*. After two more years of teaching at Leningrad, Gamow attended the International Solvay Congress on Physics in Brussels and decided not to return to the U.S.S.R. In 1934 he accepted a professorship at George Washington University; in 1956 he became professor of physics at the University of Colorado. After Gamow came to the U. S., his interests shifted from pure nuclear physics to its applications in cosmology, and later to fundamental problems of biology. A prolific popularizer of science, Gamow was awarded the Kalinga Prize in 1956 for his interpretation of science for the layman. He has published a dozen books in 23 languages, with three more going to press.

Bibliography

THE CREATION OF THE UNIVERSE. George Gamow. The Viking Press, 1952.

MODERN COSMOLOGY. George Gamow in *Scientific American*, Vol. 190, No. 3, pages 54-63; March, 1954.

THE PRIMEVAL ATOM. Georges Lemaitre. D. Van Nostrand Company, Inc., 1950.

SCIENTIFIC AMERICAN January 1958, Vol. 198, No. 1, pp. 51-57 OFFPRINT 212

THE PRINCIPLE OF UNCERTAINTY

by George Gamow

This rule of modern physics, which states that events at the atomic level cannot be observed with certainty, helps resolve the paradox that particles sometimes behave like waves and waves like particles.

It may seem a paradox that one of the cornerstones of modern physics is something called the principle of uncertainty. The idea of indeterminacy, as a rule of science does, in fact, disturb many 20th-century philosophers. But the uncertainty principle has proved a powerful answer—so far the most fruitful—to many of the most important questions in present-day physics. It is a timely moment to review how physicists arrived at the principle and to explore its meaning.

Our starting point is the classical view of the motion of a body. The concept of the trajectory of a moving object is as old as the human mind. The prehistoric caveman who threw a stone or a spear had a mental picture of its flight which told him that if he threw it in the right direction with the right speed, it would hit his prey. This basic idea of a trajectory became the foundation of men's concepts concerning the nature of motion, and in the 18th and 19th centuries mathematicians were able, on these principles, to calculate the motions of the planets with an accuracy to within an infinitesimal fraction of 1 per cent. But at the break of the 20th century a fundamental revolution occurred in physics. Men suddenly discovered that the classical laws of mechanics and energy which worked perfectly within the realm of ordinary experience did not work so well in the realm of the great cosmos on the one hand or the interior of the atom on the other. To resolve the baffling contradictions that turned up, Albert Einstein created a new relativistic view of space, time and motion, while explorers of the atom produced the quantum theory. The new ideas were so strange that they contradicted common sense, just as the notion that the earth was round had violated common sense in an earlier day when men thought the earth must be flat as a pancake because their own backyards and the land and oceans as far as they could see appeared flat. But like Magellan, whose voyage around the world proved that the earth was a globe, physicists in the 20th century soon brought forth proof that the new theories were a better description of physical nature than the comfortable old classical ideas.

In the atomic realm, two discoveries especially confounded common sense and common experience. One was the behavior of light. As a result of work by Max Planck, Einstein and others, it developed that light was made up of discrete packets of energy, named photons. The energy of the photons varied with the frequency (or wavelength) of the light: it could be stated precisely as $h\nu$, h standing for Planck's constant and ν for the frequency. Matter emitted and absorbed light only in certain definite quanta (photons). Because even dim light consists of billions of photons, we cannot detect its "graininess." But the existence of photons, as well as their obedience to the rule that their energy depends on frequency, was elegantly proved by Einstein's analysis of the photoelectric effect (ejection of electrons from a metal surface by light) and by Arthur Holly Compton's discovery of the "Compton effect" (changes in the frequency of X-rays when they lose energy in collisions with electrons).

Waves or Particles?

The proof of the existence of photons placed the classical theory of light in a very awkward position. Light was supposed to be made up of waves, and its properties of interference and diffraction showed that it did behave like waves. Yet here it was behaving like particles! To resolve this crisis, physicists had to accept the bizarre notion that light had the nature, at one and the same time, both of waves and of particles. They tried to picture it as a stream of photons which was given a wave motion by some kind of guiding "field." The picture admittedly was unsatisfactory, but it was the best they could do.

If this flew in the face of common sense, there was worse to come. It next turned out that particles behave like waves! The embarrassment grew out of Niels Bohr's famous model of the atom. He pictured the electrons as whirling around the nucleus in certain prescribed orbits: that is, they were permitted to possess only certain quanta of mechanical energy. When an electron dropped from one orbit to another, it gave up a quantum of energy in the form of light—a photon.

It was not clear what kept the electrons in their orbits or why they should occupy precisely the orbits that they did, but Louis de Broglie of France soon came forward with an answer. He suggested that the electrons were guided by waves that accompanied their motion. The nature of these "waves" was, of course, mysterious, but de Broglie found striking mathematical support for his idea. Bohr had noted that if his model was correct, the distances of the successive electron orbits from the nucleus of the atoms must be in the ratio of the squares of whole numbers—that is, 1, 4, 9, 16 and so on. This indicated the relative lengths of the circular orbits. If the motions of the electrons were guided by waves, then obviously in each orbit the length of the waves had to be such that some whole number of waves fitted exactly into the length of the orbit [*see diagram on page 89*]. Taking the hydrogen atom as a simple case, de Broglie calculated that the postulated pilot waves would fit in Bohr's orbits if the

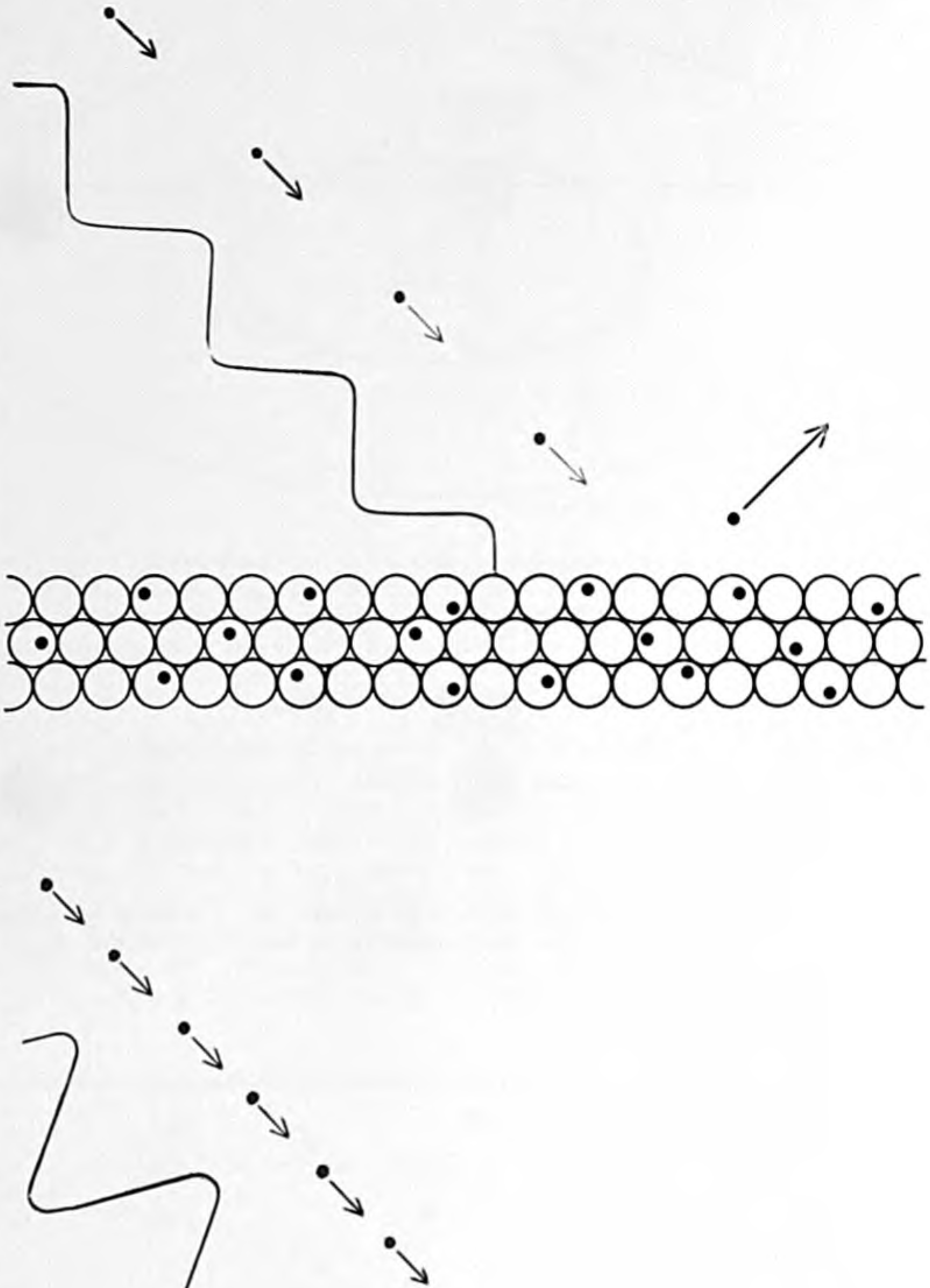
length of the wave in each case was precisely equal to Planck's constant divided by the mass and the velocity of the electron in that orbit.

From this basic idea of de Broglie, developed further by Erwin Schrödinger, came the system called "wave mechanics," a powerful theoretical tool in explaining the dynamics of the atom. Very shortly, however, the existence of de Broglie's pilot waves was demonstrated by a proof more tangible than mathematical consistency. C. J. Davisson and L. H. Germer of the Bell Telephone Laboratories performed their historic experiment showing that a beam of electrons reflected from a crystal produces a diffraction pattern—the acid test of wave motion [see "Davisson and Germer," by Karl K. Darrow; SCIENTIFIC AMERICAN, May, 1948]. And the "wavelength" of the electrons agreed exactly with de Broglie's formula.

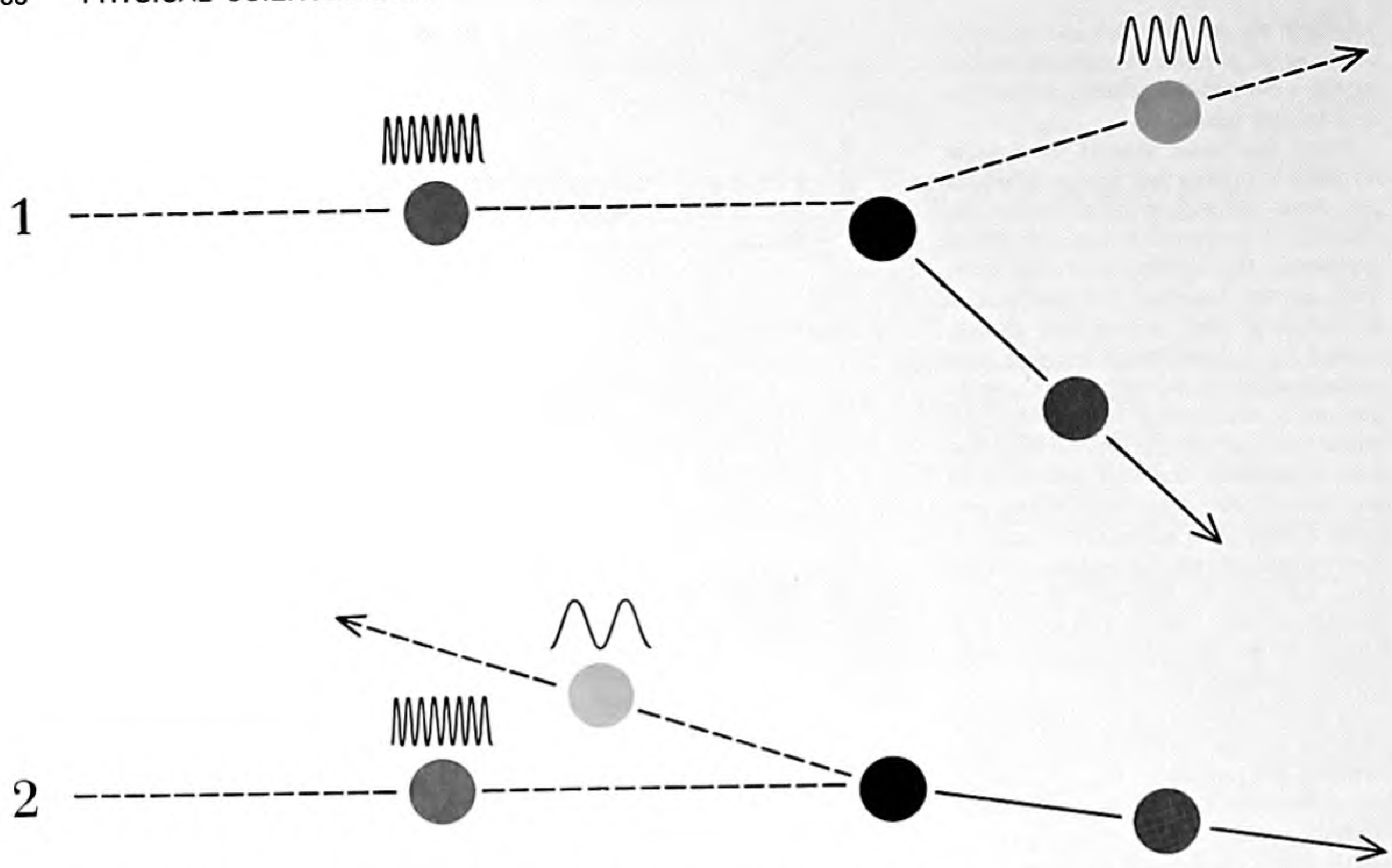
So the distinction between waves and particles all but vanished. Light waves behaved like particles, and particles behaved like waves. Common sense reeled. Whereas in the good old classical physics waves were waves and particles were particles, now one had to deal with waves possessing properties of particles and particles possessing properties of waves. Classical lines of thought could not cope with such a paradox. It was at this point that Werner Heisenberg entered to rescue common sense from total confusion. His solution was the principle of uncertainty.

A Thought Experiment

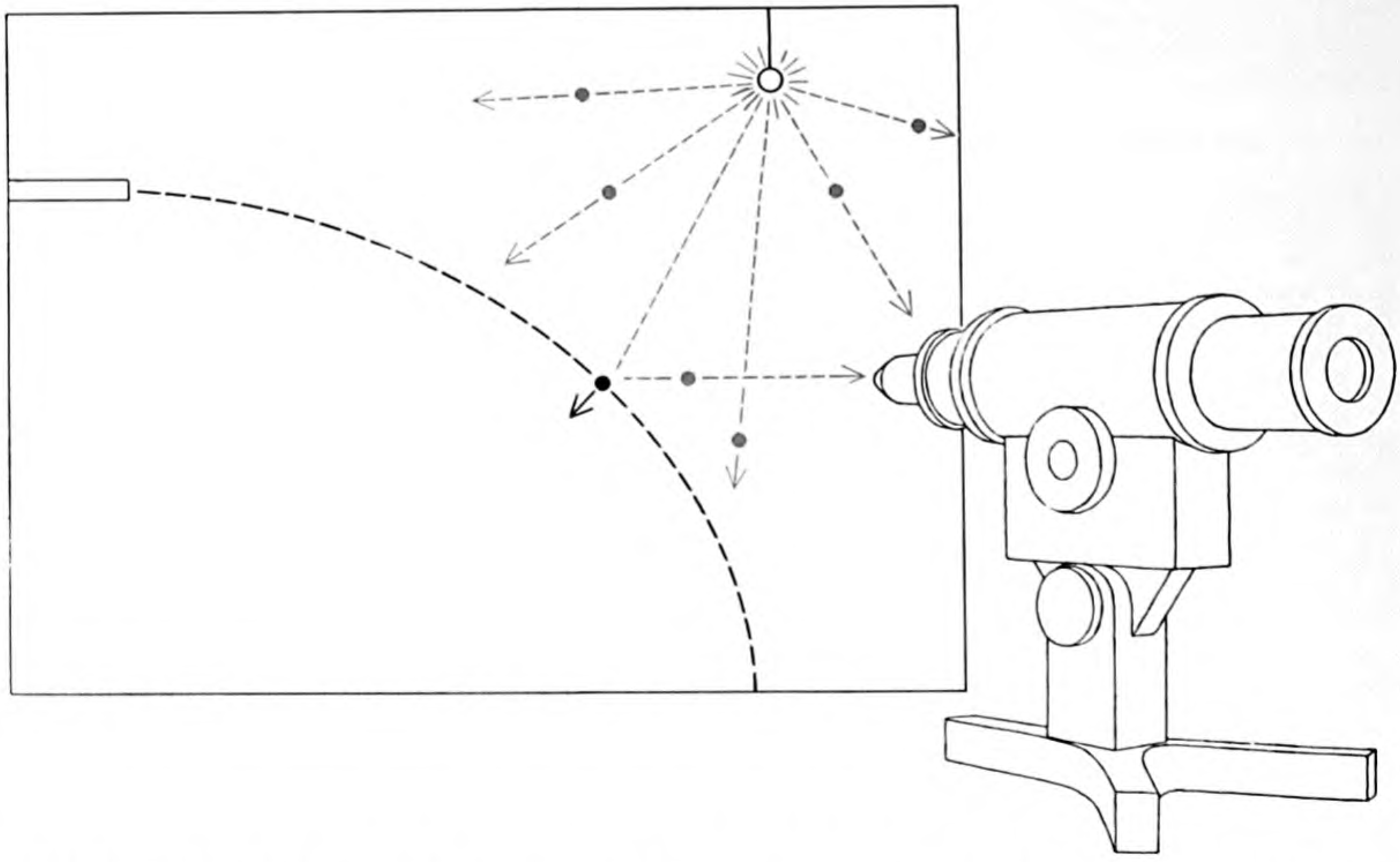
Heisenberg went back to the root of the trouble: the attempt to apply ordinary rules and methods of observation to phenomena on the atomic scale. In the world of everyday experience we can observe any phenomenon and measure its properties without influencing the phenomenon in question to any significant extent. To be sure, if we try to measure the temperature of a demitasse with a bathtub thermometer, the instrument will absorb so much heat from the coffee that it will change the coffee's temperature substantially. But with a small chemical thermometer we may get a sufficiently accurate reading. We can measure the temperature of an object as small as a living cell with a miniature thermocouple, which has almost negligible heat capacity. But in the atomic world we can never overlook the disturbance caused by the introduction of the measuring apparatus. The energies on this scale are so small that even the most



PHOTOELECTRIC EFFECT demonstrates that light waves also behave like particles. The circles in this diagram are the atoms of a metal. The black dots within the circles are electrons which can be expelled from the metal by light. The wavy lines are light waves; the colored dots, their associated particles or quanta. In the top picture light expels a single electron from the metal. In the bottom picture light of the same wavelength but greater intensity expels two electrons, each of which has the same energy as that of the first electron. This is explained on the basis that although there are more quanta associated with the second light wave, each quantum has the same energy as each quantum associated with the first light wave. On the basis of classical physics, which does not include the quantum concept, the wave of greater intensity or amplitude should expel electrons of greater energy.



COMPTON EFFECT shows that X-ray quanta (*colored balls*) deflected by electrons (*black balls*) lose energy according to their angle of deflection. That is, a quantum deflected at a large angle (1) retains more energy than one deflected at a small angle (2).



THOUGHT EXPERIMENT of Werner Heisenberg imagined a microscope which would detect a single electron (*black dot*) by means of single quanta. This illustrated his principle that phenomena at the atomic level cannot be observed without changing them.

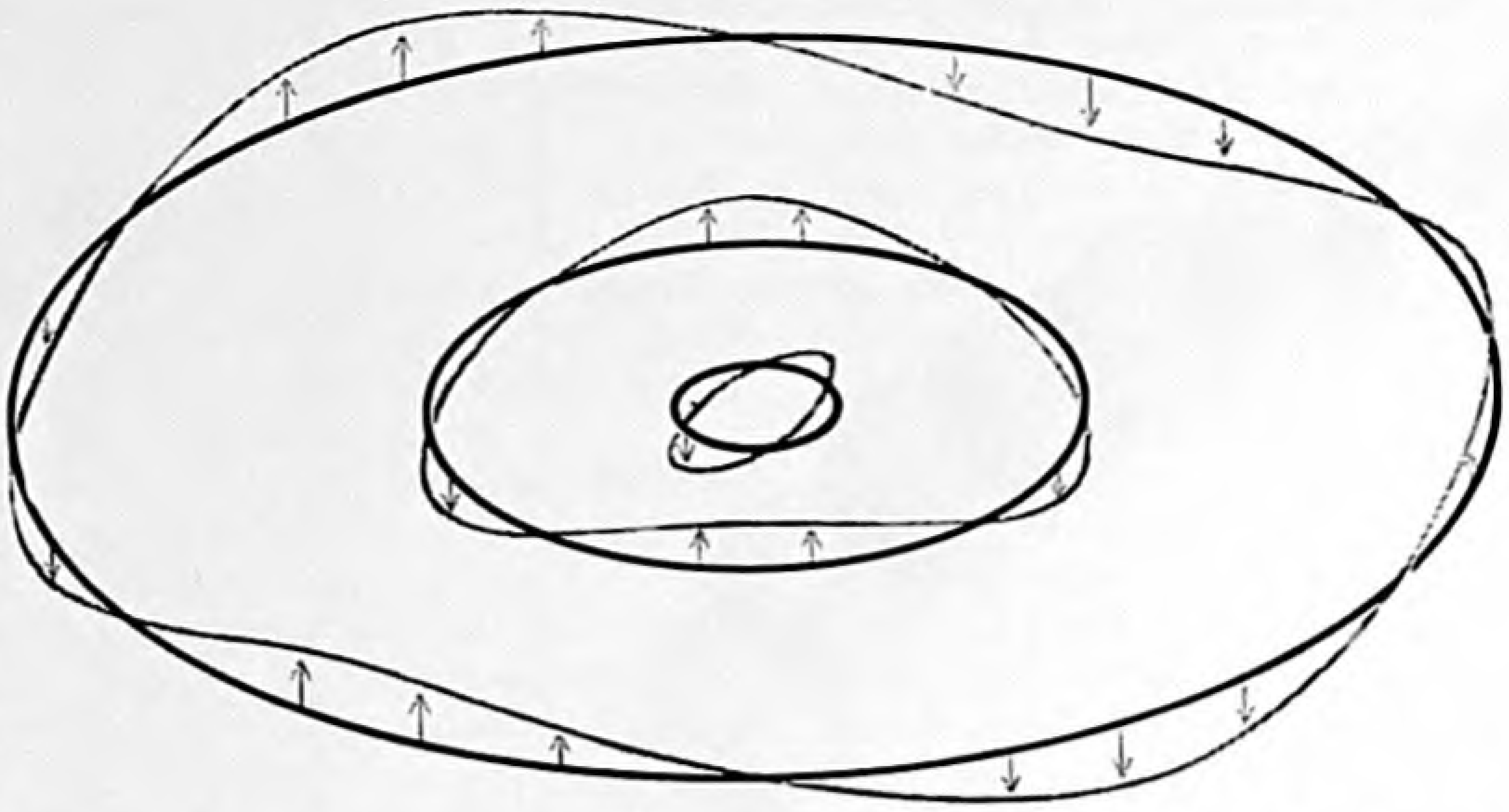
gently performed measurement may result in substantial disturbances of the phenomenon under observation, and we cannot guarantee that the results of measurements actually describe what would have happened in the absence of the measuring devices. The observer and his instruments become an integral part of the phenomenon under investigation. Even in principle there is no such thing as a physical phenomenon *per se*. In all cases there is an absolutely unavoidable interaction between the observer and the phenomenon.

Heisenberg illustrated this by a detailed consideration of the problem of trying to track the motion of a material particle. In the gross world we can follow the flight of a ping-pong ball without affecting its trajectory one iota. We know that light exerts pressure on the ball, but we do not have to play ping-pong in a dark room (assuming it were possible), because the pressure of light is much too small to make any difference in the ball's flight. But substitute an electron for the ping-pong ball, and the situation becomes quite different. Heisenberg examined the situation with a *Gedankenexperiment* ("thought experiment"), a device which was first used by Einstein in his discussion of the theory of relativity

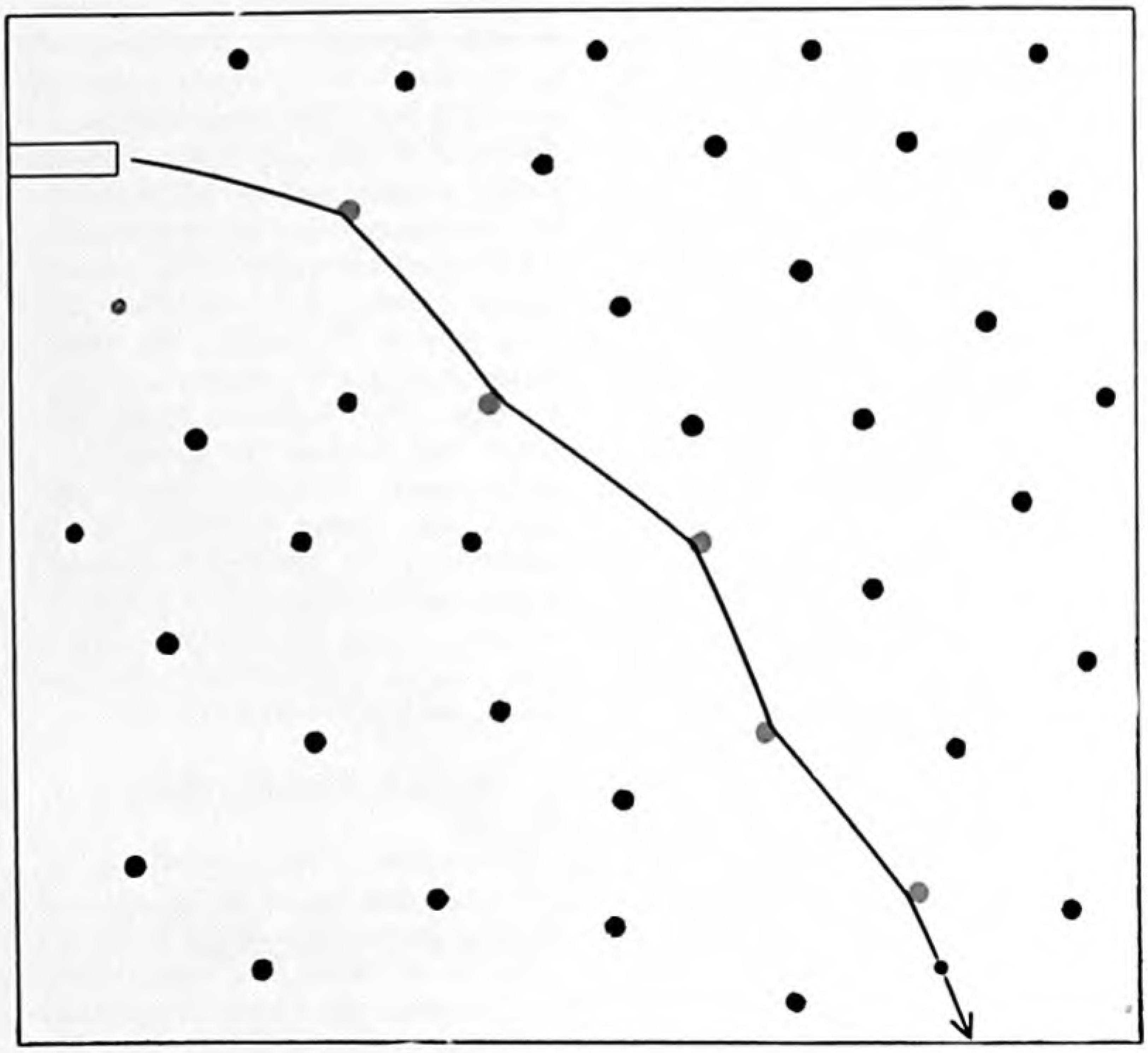
In such an exercise the experimenter is allowed an "ideal workshop" in which he can make any kind of instrument or gadget—provided that its design and functioning do not contradict basic laws of physics. For example, he can have a rocket that moves with almost the speed of light, but not more than the speed of light; or he may use a light source which emits just a single photon, but not half a photon. Heisenberg equipped himself with an ideal setup for observing the flight of an electron. He imagined an electron gun which could shoot a single electron horizontally in a completely evacuated chamber—barren of even a single air molecule! His light came from an ideal source which could emit photons of any desired wavelength and in any desired number. And he could watch the movement of the electron in the chamber through an ideal microscope which could be tuned at will over the whole range of the spectrum, from the longest radio waves to the shortest gamma rays.

The Errant Electron

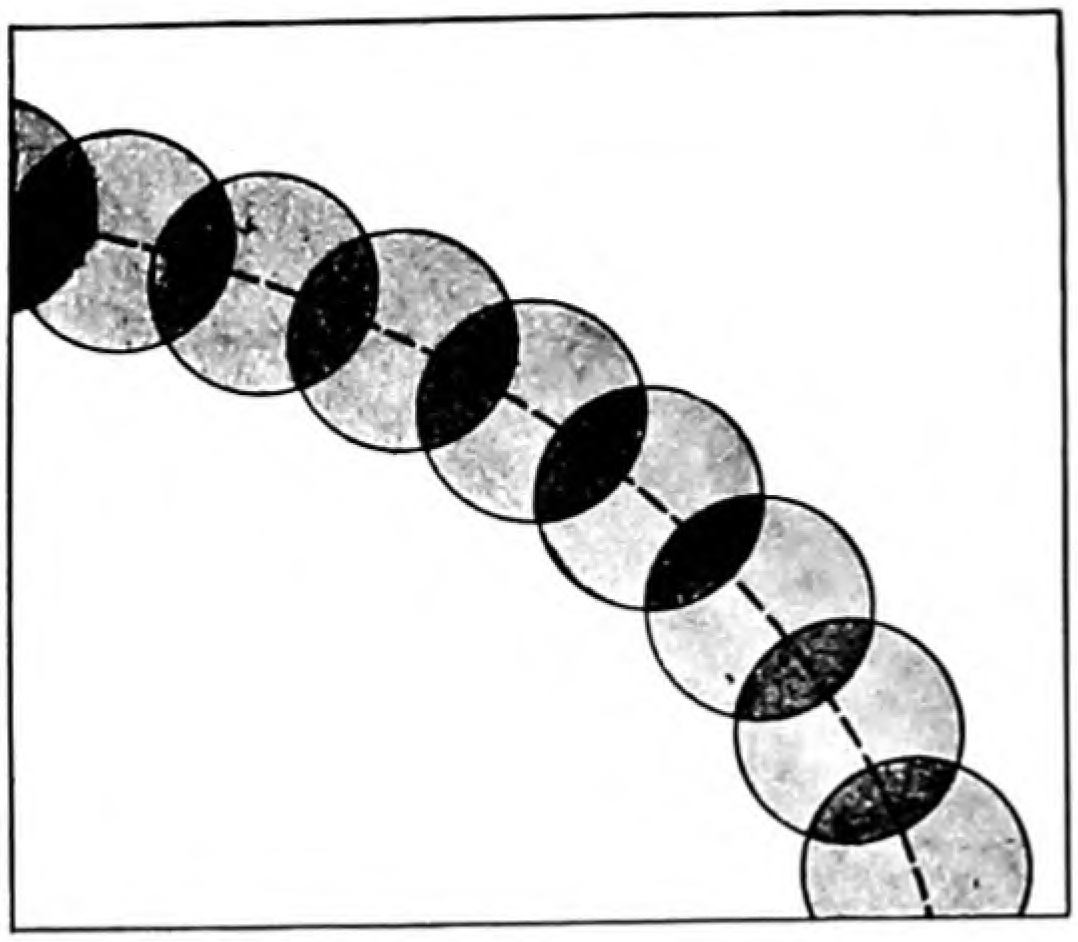
What will happen when an electron is fired in the chamber? According to our classical textbooks on mechanics, the particle should follow a trajectory known as a parabola. But actually, the moment a photon strikes it the electron will recoil and change its velocity. Observing the particle at successive points in its motion, we shall find it taking a zigzag course because of the photon impacts. Let us, then, since we have an ideally flexible instrument, minimize the impacts by reducing the photons' energy, which we can do by using light of lower frequency. In fact, by going to the limit of infinitely low frequency (which is possible in our apparatus) we can make the disturbance of the electron's motion as small as we wish. But here comes a new difficulty. The longer the wavelength of the light, the less able we are to define the object, because of the dif-

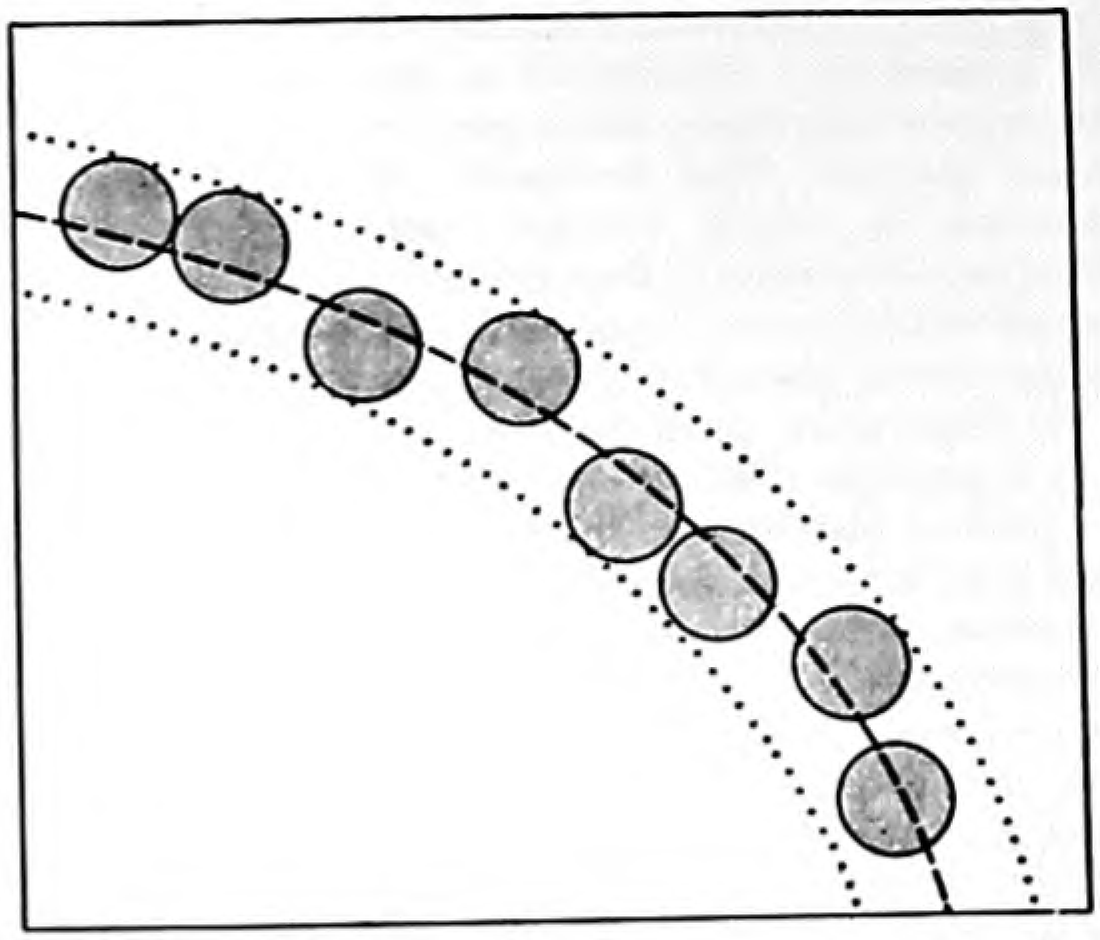


ELECTRON WAVES fitted into discrete orbits were suggested by Louis de Broglie as the explanation of the discrete energy levels of the atom. In this schematic diagram the inner orbit consists of one wave; the second orbit, of two waves; the outer orbit, of three waves. The radii of the orbits would be in the ratio of those numbers squared: one, four and nine.



SMALL INDICATORS (*larger dots*) could be used to trace the path of a particle fired from a gun (*upper left*). But the smaller the indicators, the more they disturb the path.





TRAJECTORY OF A MOVING PARTICLE is schematically observed with quanta of three different wavelengths. The trajectory of the unobserved particle is indicated by the broken line in each drawing. At left the quanta have low energy and long wavelength; they disturb the trajectory of the particle very little but only hazily indicate its path. In center the quanta have medium energy and medium wavelength; they disturb the trajectory somewhat but narrow the indication of its path. At right the quanta have high energy

fraction effect. So we can no longer find the exact position of the electron at any given instant. Heisenberg showed that the combined uncertainty in position and in velocity (*i.e.*, the product of the two uncertainties) can never be smaller than Planck's constant divided by the mass of the particle.

So with very short waves we can define the positions of a moving particle sharply but will interfere greatly with its velocity, while with very long waves we can determine its undisturbed velocity but become very uncertain about its positions. Now we can choose a middle ground between these uncertainties. If we use some optimal intermediate wavelength of light, we will disturb the particle's trajectory only moderately and still be able to define its path to a fairly close approximation [*see diagrams at top of these two pages*]. The observed path, expressed in classical terms, will not be a sharp line, but at least it will be confined within a band. Describing the trajectory of an electron in this way gives us no difficulty in a case such as a television picture tube, where the "thickness" of the electron's path to the screen is very much smaller than the diameter of the spot formed on the screen by the electron beam. Here we can represent the electron's trajectory satisfactorily by a line. But we cannot describe the orbit of an electron inside an atom in the same terms. The band of uncertainty is about as wide as the distance of the orbit from the nucleus!

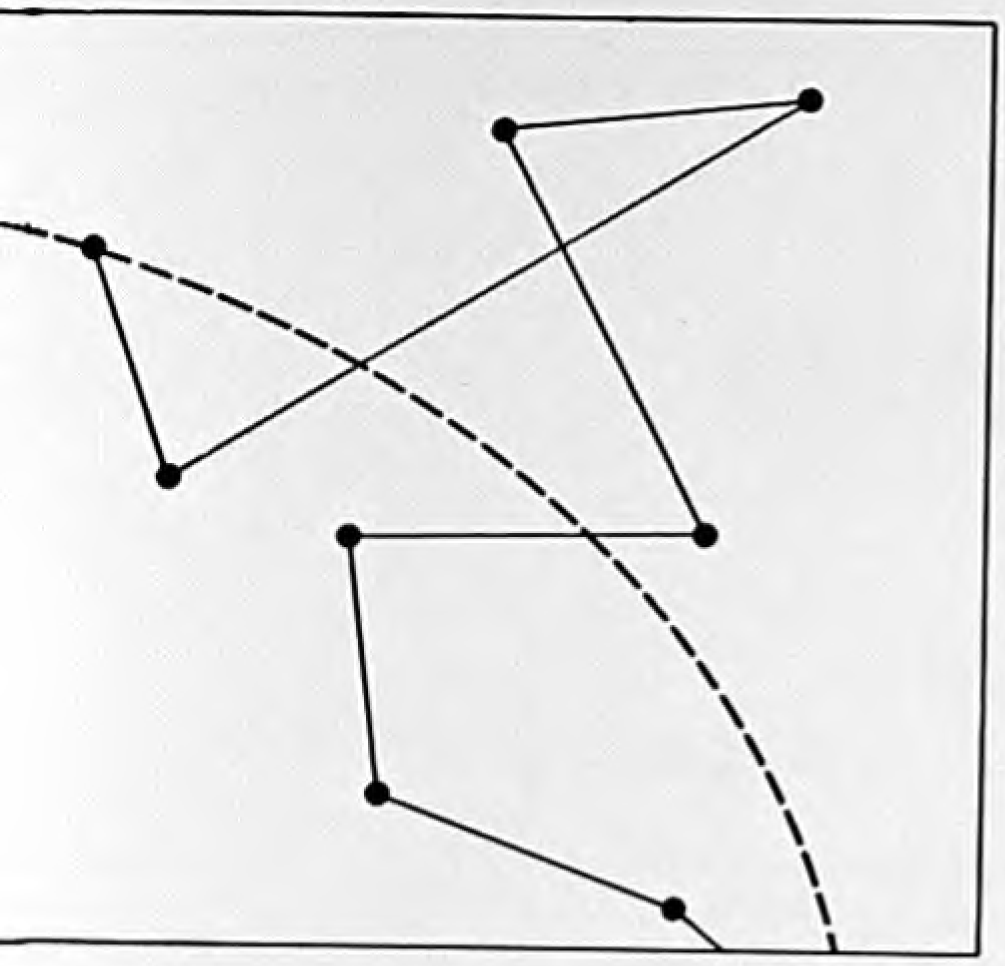
Suppose we give up the attempt to track a moving particle with light and try the cloud-chamber method instead. In our hypothetical workshop we build an ideal "cloud chamber" which is completely evacuated of material particles but is filled with very tiny imaginary "indicators" that become "activated" whenever an electron passes close by. The activated indicators would show the track of the moving particle just as water droplets do in a real cloud chamber [*see diagram at bottom of page 89*].

Classical mechanics would say that in principle the indicators could be made small enough and delicately responsive enough so that they would subtract no significant amount of energy from the moving particle and we could observe its trajectory with any desired precision. But quantum mechanics finds a fundamental objection to this procedure. One of its rules is: The smaller the system, the larger its quanta (minimum amount) of energy. Thus as the size of the "indicators" was reduced (for more precise measurement of the electron's positions), they would take more energy from the passing particle. The situation is quite analogous to the fatal difficulty in trying to track a particle by means of light, and we again arrive at the same relation for the uncertainties.

Particles Guided by Waves

Where does all this leave us? Heisenberg concluded that at the atomic level we must give up the notion of the trajectory of an object as a mathematical (*i.e.*, infinitely thin) line. This concept is accurate enough when we deal with phenomena in the realm of ordinary experience, where we can think of a moving object as held in its path by a kind of railroad track. But in the small world of photons and the atom, individual motions and events are not so firmly predetermined. Photons and material particles such as electrons and protons move over a range under the guidance of waves. The important point is that the guidance is performed in a probabilistic rather than a strictly deterministic way. We can measure only the ***probability*** that a photon will strike a given point on a screen, or that a material particle will be found in a given place at a given instant.

I must make clear that the word "probability" here is used in a rather different sense from the way it is usually understood in classical physics and everyday life. When we say in a game of poker that there is a certain probability of drawing a royal flush, we mean only that we have to estimate the chances because we do not know the arrangement of the cards in the pack. If we knew exactly how the cards were stacked, we could predict definitely whether we would get a royal flush or not. Classical physics assumed that the same was true of a problem such as the behavior of a gas: its behavior had to be described on the basis of statistical probability only because of incomplete knowledge—if we were given the positions and velocities of all the particles, we could predict events within the gas in full detail. The uncertainty principle cuts the ground from under that idea. We cannot predict the motions of individual particles because we can never know the initial conditions exactly in the



and short wavelength; each quantum sharply locates the particle but completely disturbs its trajectory. Thus the trajectory can at the most be only roughly approximated.

first place. It is impossible *in principle* to obtain an exact measurement of both the position and the velocity of a particle on the atomic scale.

A look at Heisenberg's formula for measuring uncertainty shows why we can disregard the principle of uncertainty and safely trust the good old principle of determinism when we deal with matter on the macroscopic scale. The uncertainty, as I have mentioned, is equal to Planck's constant h divided by the mass of the particle. Planck's constant is an extremely small quantity: its numerical value amounts to only about 10^{-27} in centimeter-gram-second units. When we consider a particle weighing as much as one milligram, we can in principle simultaneously determine its position within a trillionth of a centimeter and its velocity within a trillionth of a centimeter per second—or 30 microns per century!

Heisenberg's principle was developed by Bohr into a new philosophy of physics. It called for a profound change in our ideas about the material world—ideas that we acquire in ordinary experience from early childhood. But it allowed many puzzles of atomic physics to make sense. Above all, it extricated us from the wave-particle paradox. The uncertainty principle showed that the wave and particle ideas are mutually complementary ways of describing nature.

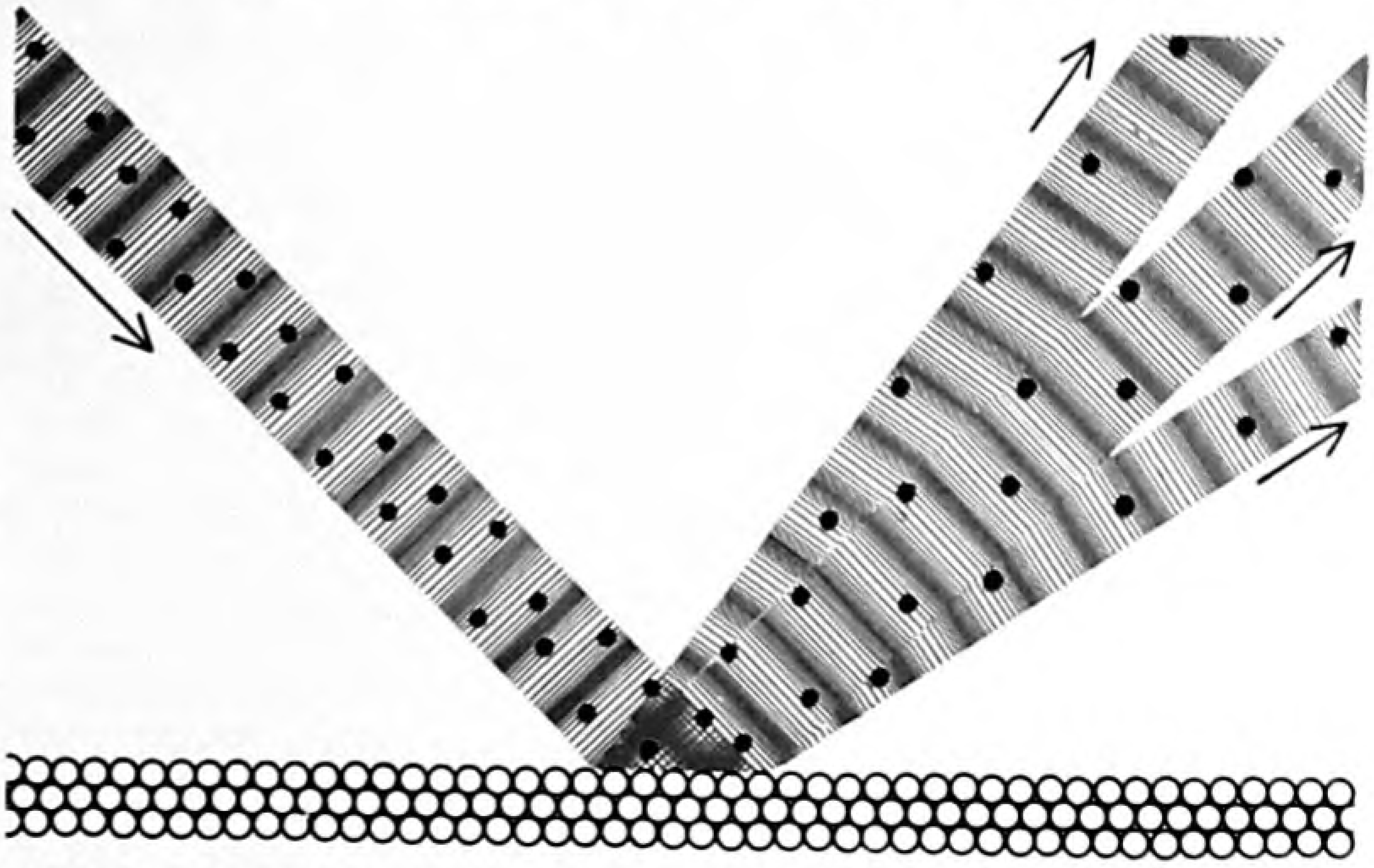
Many physicists readily accepted the new view. Others did not like it at all. To the latter group belonged Albert Einstein. His philosophical convictions about determinism did not permit him to elevate uncertainty to a principle. And just as skeptics were trying to find contradictions in his theory of relativity, Einstein tried to discover contradictions in the uncertainty principle of quantum physics. However, his efforts led only to strengthening of the principle of uncertainty. This is interestingly illustrated by an incident that took place at the sixth International Solvay Congress on Physics, in Brussels in 1930.

In a discussion at which Bohr was present, Einstein performed a "thought experiment." Arguing that time was a fourth coordinate of space-time and that energy was a fourth component of momentum (mass times velocity), he said that Heisenberg's uncertainty equation implied that the uncertainty in time was related to the uncertainty in energy, the product of the two being at least equal to Planck's constant h. Einstein set out to prove that this was not the case—that the time and the energy could be determined without any uncertainty. Consider, he said, an ideal box, lined with perfect mirrors, which could hold radiant energy indefinitely. Weigh the box. Then at a chosen instant some time later a clockwork, preset like a time bomb, will open an ideal shutter to release just one photon. Now weigh the box again. The change of mass tells the energy of the emitted photon. In this manner, said Einstein, one could measure the energy emitted and the time it was released with any desired precision, in contradiction to the uncertainty principle.

The next morning, after an almost sleepless night, Bohr delivered a mortal blow to Einstein's disproof. He offered a counter thought experiment with an ideal apparatus of his own (which I later actually built in wood and metal, as Bohr's student, for his use in lectures on the subject). Bohr attacked the question of weighing Einstein's box. A spring scale equipped with a pointer recording the weight on a vertical column placed alongside is, he said, as good as any. Now since the box must move vertically with a change in its weight, there will be an uncertainty in its vertical velocity and therefore an uncertainty in its height above the table, Bohr pointed out. Furthermore, the uncertainty about its elevation above the earth's surface will result in an uncertainty in the rate of the clock, for according to the theory of relativity the rate depends on the clock's relative position. Bohr proceeded to show that the uncertainties of time and of the change in the box's mass would indeed have the relation which Einstein had tried to disprove.

Einstein, bitten by his own argument, had to agree that Heisenberg's concept was free of internal contradictions, but to the very end of his life he refused to accept the uncertainty principle and remained hopeful that physics would some day return to the deterministic point of view.

During the past decade the validity of the uncertainty principle has been argued voluminously, both by writers who understand the problems at issue and by writers who do not. Up to the present this so-called "Copenhagen interpretation of the quantum theory" has stood its ground. In my opinion and in the opinion of many other theoretical physicists, the uncertainty principle will stand its ground indefinitely.



ELECTRON WAVES WERE DISCOVERED by C. J. Davisson and L. H. Germer when they observed that a beam of electrons (*left*) was deflected at several discrete angles (*right*) by a crystal lattice (*bottom*). This is characteristic of diffraction, a wave phenomenon.

The Author

GEORGE GAMOW was present at the birth of "the philosophy of uncertainty." As a graduate student of physics at the University of Leningrad in 1928, he found himself in hot water because "the research topic proposed by my professor was so boring that the work hardly made any progress." That summer Gamow took some courses at the University of Göttingen. While there he devised a quantum theory of radioactivity which shed new light on the structure of the atomic nucleus. Instead of returning to Leningrad he accepted an invitation from Niels Bohr, who had learned of his work, to spend a year at the University of Copenhagen on a Fellowship provided by the Carlsberg brewing firm. The following year he worked with Ernest Rutherford at the University of Cambridge on a Rockefeller fellowship. At Cambridge he wrote his first book, entitled *Constitution of Atomic Nuclei and Radioactivity*. After two more years of teaching at Leningrad, Gamow attended the International Solvay Congress on Physics in Brussels and decided not to return to the U.S.S.R. In 1934 he accepted a professorship at George Washington University; in 1956 he became professor of physics at the University of Colorado. After Gamow came to the U.S., his interests shifted from pure nuclear physics to its applications in cosmology, and later to fundamental problems of biology.

Bibliography

ADVANTAGES AND DISADVANTAGES OF VARIOUS INTERPRETATIONS OF THE QUANTUM THEORY. Henry Margenau in *Physics Today*, Vol. 7, No. 6, pages 6-13; October, 1954.

DISCUSSION WITH EINSTEIN ON EPISTEMOLOGICAL PROBLEMS IN ATOMIC PHYSICS. Niels Bohr in *The Library of Living Philosophers*, Vol. 7, pages 199-241; 1949.

NUCLEAR PHYSICS. W. Heisenberg. Philosophical Library, 1953.

PHYSICS AND MICROPHYSICS. Louis de Broglie. Hutchinson's Scientific and Technical Publications, 1955.

SCIENTIFIC AMERICAN July 1957, Vol. 197, No. 1, pp. 72-88

ELEMENTARY PARTICLES

by Murray Gell-Mann and E. P. Rosenbaum

An account of the abstract theoretical ideas which physicists use to help them understand the material world. These ideas begin to show some order in the jumble of subatomic particles.

There is no excellent beauty that hath not some strangeness in the proportion.
—Francis Bacon

This aphorism quoted from Bacon is doubtless true of science as well as art. But is too much strangeness not fatal to beauty? For years strangeness has afflicted one of the basic concerns of physics: the nature of matter. When physicists considered matter on the smallest scale, it appeared to be an arbitrary jumble of elementary particles. No simple and orderly relationship among the particles could be perceived. Now at last the picture seems to be clearing up a bit. The very word "strangeness" has passed into the vocabulary of physics, but its share "in the proportion" is being reduced to a point where the beauty of order can be seen.

The new regularity can best be appreciated against the chaotic background from which it is emerging. To begin we should go back some 30 years to one of the most triumphant periods in the history of science. The theory of the atom stood essentially complete: nearly all the properties of ordinary matter could be mathematically deduced in terms of the motions of negatively charged electrons around positively charged nuclei. Most of the problems with which physics and chemistry had grappled during the preceding centuries were in principle solved. But at that time physicists began seriously to probe the interior of the atomic nucleus.

Then their troubles began. They soon learned that the nucleus is made up of protons and neutrons, but they could not explain nuclear properties only in terms of these constituents. Indeed, we still do not know exactly what their motions are. Furthermore it turned out that when a nucleus is shattered, entirely new types of matter are created—a bewildering variety of short-lived particles which apparently do not exist within the atoms of ordinary material. Some of them were reasonably well accounted for when they turned up, but others fit nowhere in the physicist's scheme of nature. They were called "strange" particles.

The First Particles

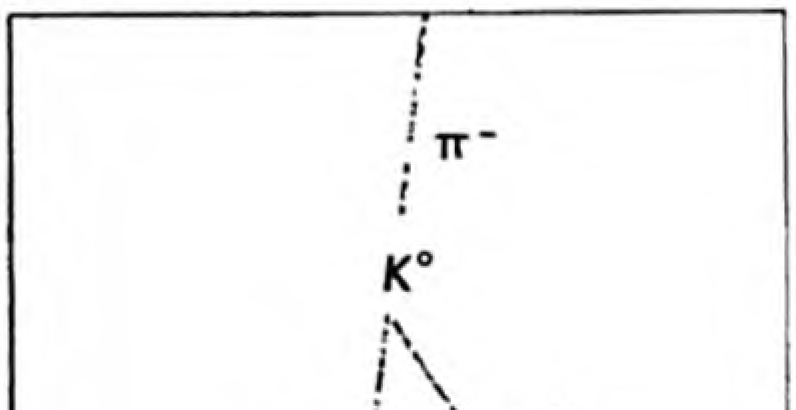
We are getting ahead of our story. We should begin in the early 1930s, when the atomic drama had only four characters: the electron, proton, neutron and photon. The first three are the building blocks of atoms—protons and neutrons in the nucleus and electrons in the space around it. The photon is the quantum unit of radiation; *i.e.*, it is the building block of the electromagnetic field.

The photon always travels with the velocity of light (denoted by the letter "c"); it can never be at rest. Because of its motion it possesses energy. It therefore also possesses mass, according to the famous relation $E = mc^2$. But the mass exists only by virtue of the motion. The electron, proton and neutron, on the contrary, can be at rest. Each has a mass when at rest and a corresponding rest energy. (When in motion, of course, they have additional energy and mass.)

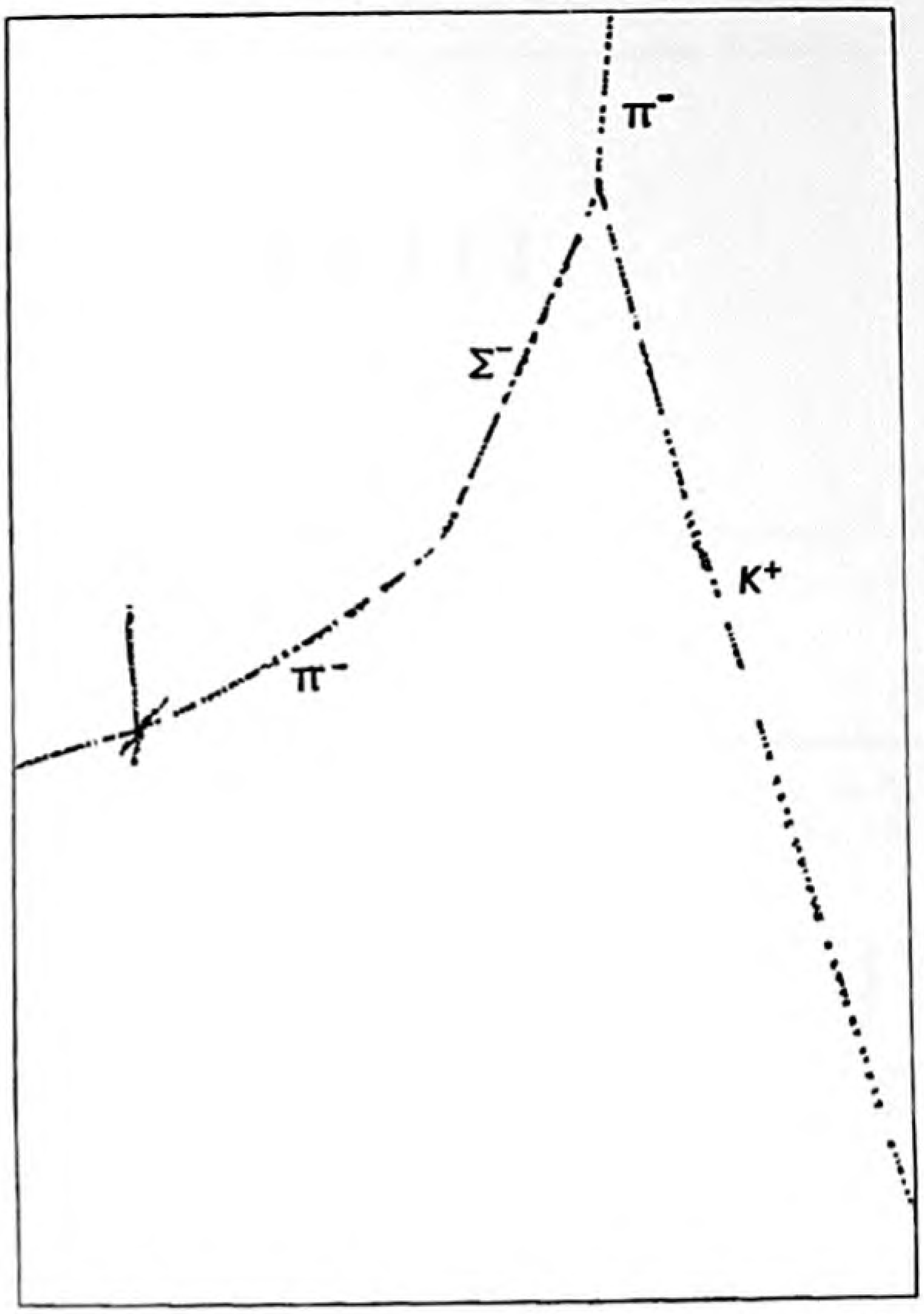
The electron is the lightest particle with any rest mass, and this mass is a basic unit in subatomic physics. The size of the electron's negative charge is likewise a basic unit of electricity. In these units the proton has a mass of about 1,836.1 and a charge of plus one; the neutron has a mass of about 1,838.6 and no charge. The photon, as we have said, has no rest mass; also it has no charge, although it is the carrier of electromagnetic energy.

All these particles spin on their axes and, if they are charged, the spin makes them tiny magnets. According to the rules of quantum theory the spin has a fixed rate characteristic of the particle. In the system of units used in quantum theory, the characteristic spin of the electron, proton and neutron is 1/2; the spin of the photon is 1.

There is a further limitation on the spinning motion of these particles. If they are magnets, they are affected by external magnetic fields. In quantum mechanics the spin axis of each particle can assume only a few fixed directions with respect to an outside field. A particle with spin 1/2 can have two positions: its axis can point with or against the field. A particle with spin 1 can have



NEUTRAL K PARTICLE, formed in a collision of a negative pion with a proton, decays into a pair of oppositely charged pions [*see photograph on page 106*].



SIGMA AND K PARTICLES are produced together when a pion hits a proton in the bubble chamber. The sigma decays to a pion and an invisible neutron. The pion then hits a carbon nucleus and makes a "star." The experiments shown here and on the preceding page were performed by J. H. Steinberger, N. P. Samios, R. J. Plano, F. R. Eisler and M. Schwartz, all of Columbia University.

three positions: its axis can go with the field, perpendicular to or against the field [*see diagram on page* 99].

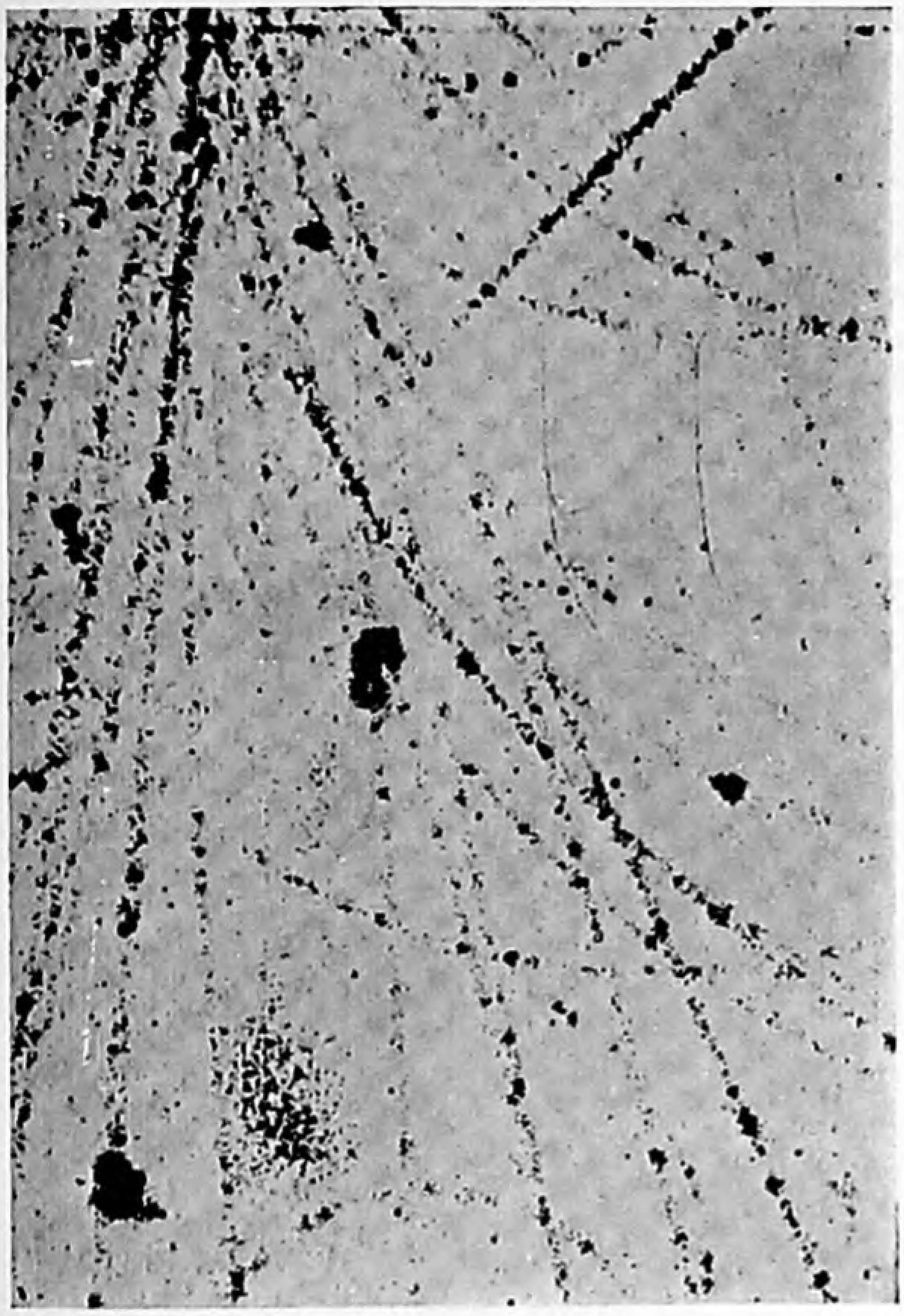
Another important property of particles, related to spin, is their "statistics." Electrons, protons and neutrons (and all other particles of spin 1/2) obey the famous exclusion principle. This says that only one particle of a kind can occupy a given quantum "state." Thus there can be only one electron at a time spinning in a particular direction and revolving in a given orbit around a nucleus. Particles which obey the exclusion principle are said to have Fermi-Dirac statistics: they are accordingly called fermions. Particles like the photon (and all other particles whose spins are whole numbers) do not obey the exclusion principle. They have Bose-Einstein statistics and are called bosons.

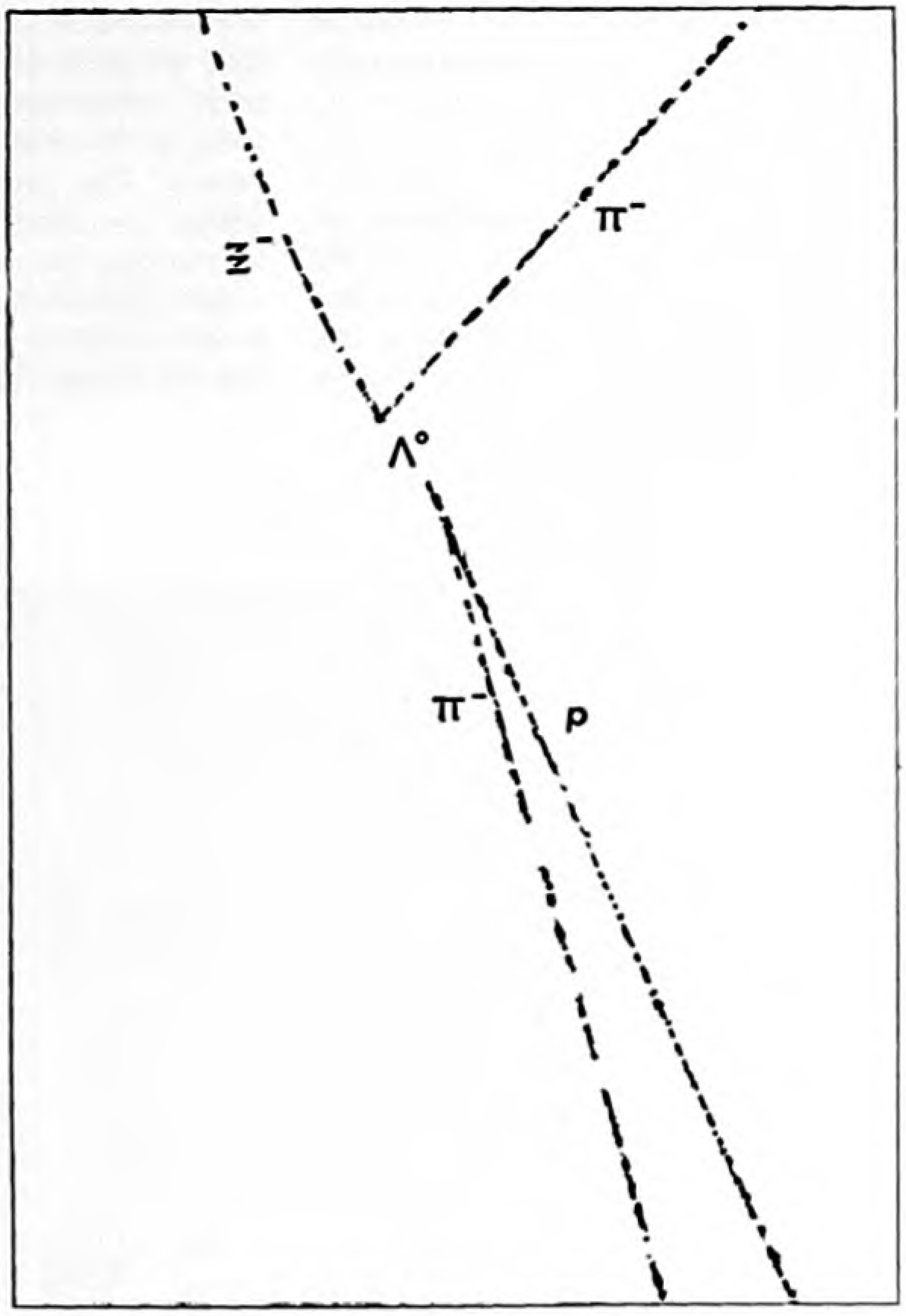
Interactions

So far we have been talking mainly about isolated particles. However, as will become increasingly evident, all particles are "coupled" to one another: when they are close together, they interact in various ways. The first such coupling to be recognized and studied was the one between the electron and the photon. It is this relationship which underlies quantum electrodynamics, the crowning achievement of atomic theory. Many physicists had a hand in the development of this theory, notably P. A. M. Dirac of England, Werner Heisenberg of Germany and Wolfgang Pauli, now in Switzerland. The theory explained the behavior of electrons in electromagnetic fields by saying that each electron continuously emits and absorbs photons. This pulsation is, so to speak, a "vital process" of the electron, and it is the means by which field and electron exert a force on each other.

We should point out that what has just been said does hardly more than name the theory. In quantum mechanics a theory is a set of mathematical relations which, given the interacting particles and the couplings between them, predicts their behavior in detail by yielding the probability of every possible reaction among the particles. Sometimes, particularly when the couplings are very strong, the mathematics turns out to be too difficult. Then the theory is not much help. In quantum electrodynamics, however, the mathematics is tractable, and this beautiful theory has successfully predicted the outcome of every fundamental atomic experiment at least as accurately as physical measurements can be made.

The basic "reaction" of quantum electrodynamics, in which electrons emit and absorb photons, is an example of what is called a virtual process. This concept, which concerns all the elementary particles, is peculiar to quantum theory. It involves an apparent violation of the law of the conservation of energy. The point is that a photon has energy; thus when a photon is spontaneously emitted by an electron, it would appear that the total energy of the system has suddenly increased. Quantum theory answers, in essence, that the photon is emitted and reabsorbed so fast that the gain in energy cannot be detected, even in principle. That is what is meant by a virtual pro-





NEGATIVE XI PARTICLE decays into a neutral lambda and a negative pion. The lambda then decays into a proton and a second negative pion. The xi was produced by the collision of a high-energy cosmic ray with a nucleus in a lead plate. This photograph, which shows tracks in a cloud chamber below the plate, was made by E. W. Cowan of the California Institute of Technology.

cess. If the photon is undetectable, the conservation of energy is effectively not violated because, according to quantum mechanics, laws deal only with observable quantities. By adding enough energy from the outside (for example, by accelerating the electron) the photons can be converted from virtual to real particles.

Virtual photons are involved in every interaction between charged bodies and electromagnetic fields. The positive proton is also considered to emit and absorb virtual photons. Here, however, the theory is not quite so successful; its predictions for protons are not as accurate as those for electrons.

In an important sense the scheme we have described so far was complete and satisfactory. Between them the electron and photon sufficed to explain all the external properties of atoms; the proton and neutron accounted for the observed charges of atomic nuclei and roughly for their masses. There was, to be sure, nothing in the theories to explain why nature had chosen just these particles as her elementary building blocks; but given that she had, they came close to being all that was needed.

Antiparticles

They did not come quite close enough. First of all, Dirac's theory of the electron predicted some additional particles [*see column II of chart on next two pages*]. It is well known that according to quantum theory a fundamental particle also has the properties of a wave. When Dirac's wave equation for the electron was solved, it yielded a negative frequency as well as a positive. Since frequency in quantum mechanics is proportional to energy, it was at first hard to see what the negative answer could mean. Dirac was able to prove that it does have a physical significance, and that it corresponds to an electron with positive charge. Furthermore, according to the theory, if a positive electron collided with a negative electron, they would annihilate each other and their mass would be converted into photons with an equivalent amount of energy. Conversely, if enough energy could be concentrated in a small volume, as in a high-speed collision between two particles, a positive and a negative electron could be created.

These remarkable predictions were not actually made (although they were implied by the theory) until Carl D. Anderson of the California Institute of Technology discovered the positron. It had the mass of an electron and a unit of positive charge; when it met with a negative electron, the two were annihilated; it could be created, together with a negative electron, in energetic collisions. The positron is called the antiparticle of the electron because it cancels out an ordinary electron.

There are similar equations for the proton and neutron, so they also have their antiparticles. These have only been detected during the past two years [see "The Antiproton," by Emilio Segrè and Clyde E. Wiegand; SCIENTIFIC AMERICAN Offprint 244]. Even the photon has an antiparticle in a mathematical sense. Here, however, the two solutions to the equation can be interpreted in the same way and the photon and antiphoton

are indistinguishable. To put it another way, the photon is its own antiparticle.

The Neutrino

The second necessary addition to the list of particles arose out of the behavior of the neutron. Inside the nucleus a neutron can live indefinitely. But when the particle is observed outside, it proves to be unstable. In an average time of about 18 minutes it spontaneously ejects a beta particle (the same thing as an electron) and turns into a proton. The proton and electron together are about 1.5 electron masses lighter than the neutron, so this amount of mass appears to be lost in the decay; it is equivalent to some 780,000 electron volts of energy. This should show up as the kinetic energy of the decay products, but in fact the proton and electron rarely have so much energy. To account for the discrepancy Pauli suggested that another particle, with zero rest mass and almost undetectable, also is formed in the decay, and that it carries off the missing energy. Enrico Fermi, who pursued the idea, named the invisible particle the neutrino. Reasoning

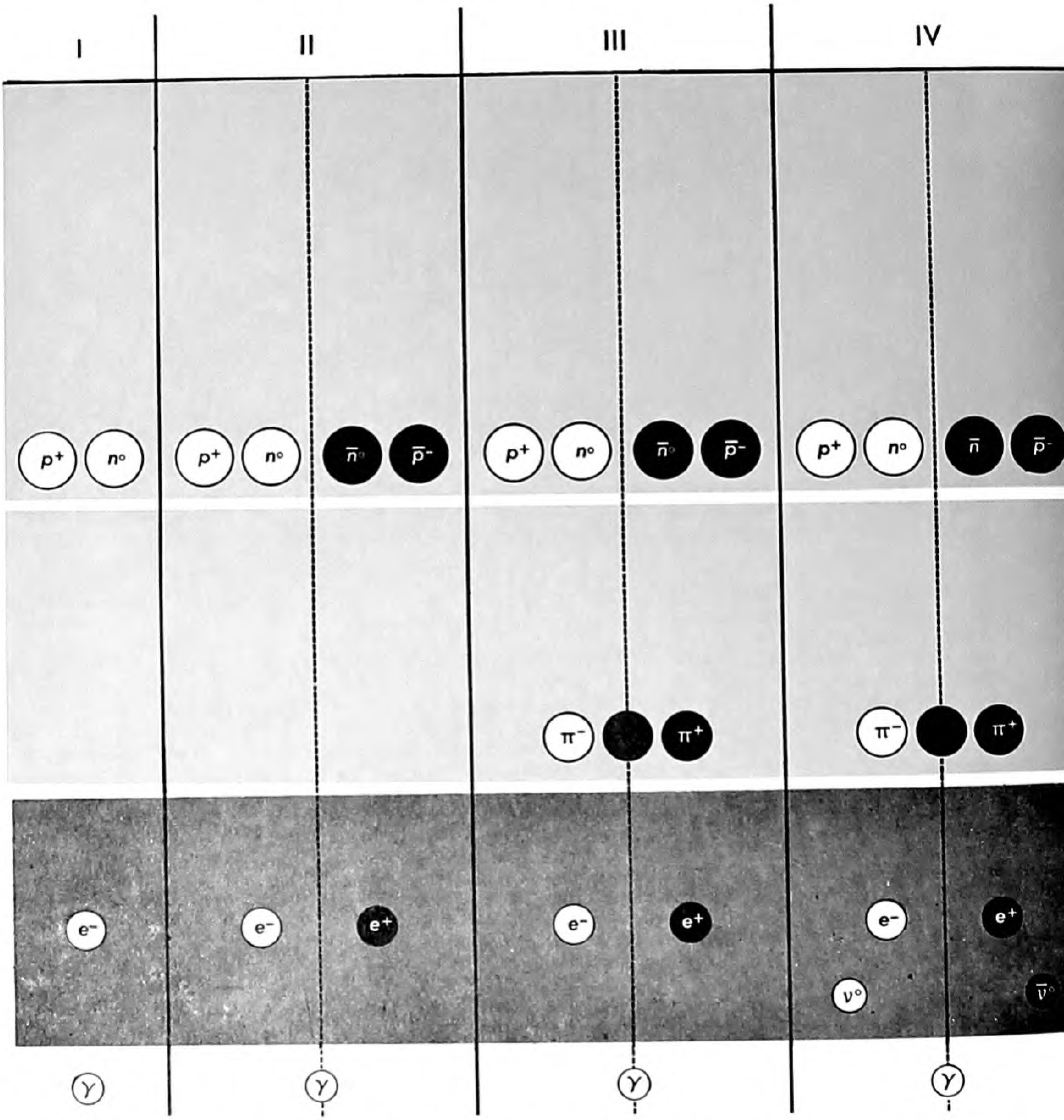
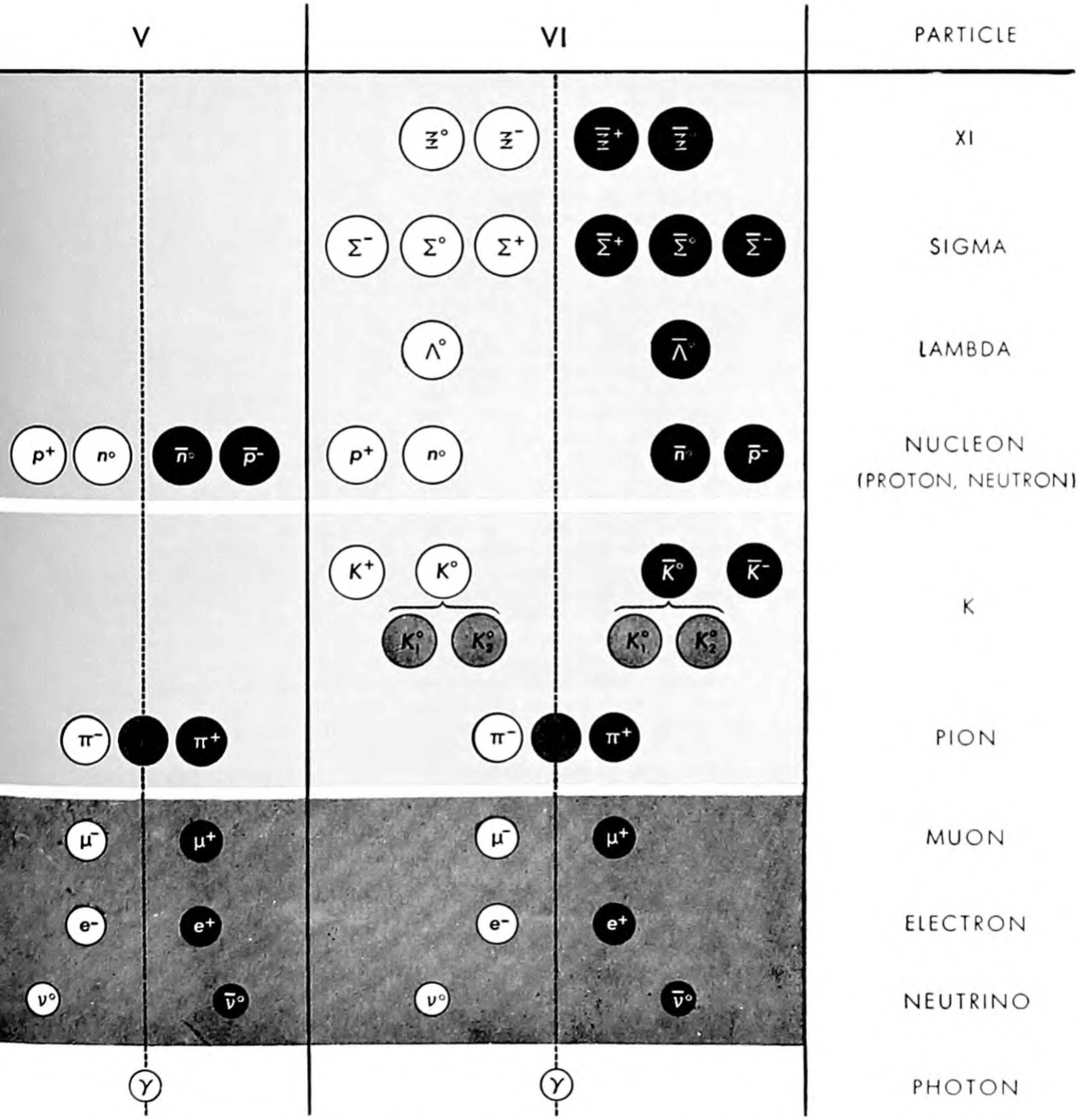


TABLE OF PARTICLES traces their increase over the past 25 years. Particles added in columns II-IV were first predicted theoretically. Those added in V and VI were discovered by experiment. "Ordinary" particles are shown as white balls, ahtiparticles as black balls. The neutral pion and the photon is each its own antiparticle. The top group comprises the heavy particles and the next

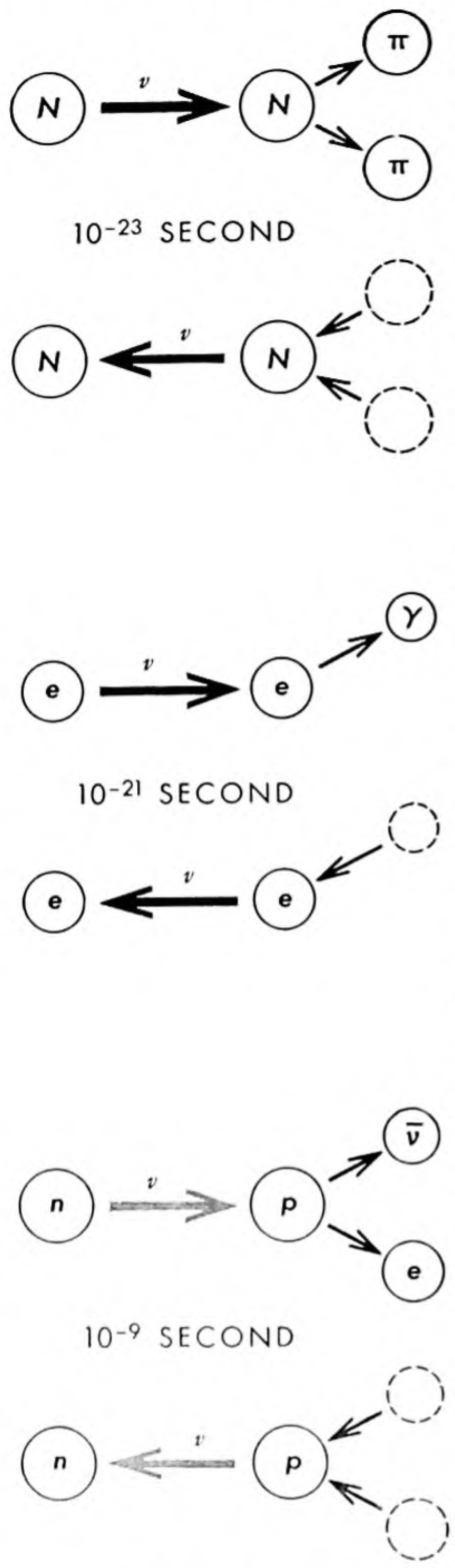
by direct analogy with Dirac's process for electrons and photons, Fermi constructed a complete theory of beta-decay. Its fundamental process is that a neutron continuously loses and regains an electron and a neutrino by virtual emission and absorption. (Strictly speaking the "neutrino" involved is actually the antineutrino.) Although Fermi's reaction was written as a virtual process, the emission or decay process can become real without the addition of outside energy because the mass lost in the decay provides the energy needed.

The Pion

The last particle to be added to the list was predicted by another analogy with the Dirac process. The problem was to describe the force that holds protons and neutrons (which may jointly be called nucleons) together in the nucleus. Since electromagnetic forces had been successfully explained in terms of the photon or field quantum, it was logical to try the same approach with nuclear forces. The Japanese physicist Hideki Yukawa took this step. He proposed that nucleons emit and absorb a



lower group the mesons. The color over both shows they are all strongly coupled. The wide band between the groups signifies that heavy particles are conserved. The third group, in gray, comprises the light particles. A bar over a symbol stands for an antiparticle, but the convention is not used for the pion, muon or electron. For explanation of K_1^0 and K_2^0 see page 107 of the text.



BASIC PROCESSES are shown schematically. In the strong Yukawa interaction (*top*) a nucleon "virtually" emits pions (*above*) and absorbs them (*below*). Time scale for the process is listed between the two reactions. In the electromagnetic interaction (*center*) an electron (or other charged particle) virtually emits and absorbs a photon. In the weak and very much slower Fermi interaction (*bottom*) a neutron virtually emits and absorbs an electron and antineutrino.

nuclear-field quantum called a meson, just as electrons emit and absorb photons. From the known properties of the nuclear force Yukawa was able to deduce some of the characteristics of the meson. The fact that the force extends only over a very short range could be shown to mean that the meson, unlike the massless photon, would have a finite rest mass. There were also various reasons to suppose that there would be both charged and neutral mesons.

Yukawa's conjecture was fully confirmed, but only after more than 10 years. The particle he predicted has been found and is now called the pi meson, or pion. It weighs about 270 electron masses, and comes in three forms: positive, negative and neutral [see "Pions," by Robert E. Marshak; SCIENTIFIC AMERICAN Offprint 226].

The emission of pions by a nucleon must of course be virtual, since the pions possess energy including energy in the form of rest mass. According to the theory the strength of the nuclear force field should depend on the number of quanta outside the emitting particle. The nuclear force is so strong that a nucleon must emit pions very frequently; there must usually be more than one outside the nucleon at the same time. In fact, the current conception of protons and neutrons is that they consist of some sort of core surrounded by a pulsating cloud of pions. As in the case of photons, if enough energy is supplied pions will materialize into real particles. Since the pion's mass is equivalent to 135 million electron volts (mev) of energy, it requires at least this amount to make one real pion.

Here we come to the end of the first part of the story. We have already accumulated a rather large number of "elementary" particles, but at least they all seem to make sense. In fact, most of them were predicted theoretically before they were actually discovered. As has already been pointed out, the neutrino has an antiparticle. The negative pion is the antiparticle of the positive pion, and *vice versa;* the neutral pion, like the photon, is its own antiparticle. Finally real pions, like the neutron, are unstable. After a very short time they decay into other particles [*see table on page 100*].

Twelve Particles

We might call the ideas we have sketched so far the dozen-particle theory of matter [*see column IV in the chart on the preceding two pages*]. As we have said, it is a fine theory for explaining the properties of atoms. It is rather crude in its attempt to account for the inner workings of the nucleus, but it does explain them in a general way. And in any event it makes a good case for each of the particles. They all have an explicit role to play and they emerge naturally from the theory.

What is more, the 12 fall into four well-defined groups: (1) heavy particles, consisting of the nucleons (proton and neutron) and their antiparticles; (2) mesons, or particles of intermediate weight; (3) light particles, consisting of the electron and neutrino and their antiparticles; and (4), in a class of its own, the photon. We may also note that the heavy and light particles, which are the "ordinary" constituents of matter, have spin 1/2 and are fermions [*see table on page 100*]. The mesons and the photon, which are field quanta, have spin zero and are bosons. The groups are interconnected by three basic reactions. The Yukawa process connects heavy particles with mesons, the Dirac process connects light particles with photons, and the Fermi process connects heavy particles with light particles [*see chart on this page*].

Of course the particles also behave according to the more general laws of physics. They obey the conservation of energy, and of linear and angular momentum. They also obey the conservation of charge. So far as we know the net amount of electric charge in the universe never changes. When charged particles are created out of energy, they can be created only in particle-antiparticle pairs, with each new positive charge offset by a negative. And in every particle reaction the net charge of the bodies entering the reaction must equal that of the products.

Another conservation principle arises out of the evident stability of nuclear matter. All the experimental evidence indicates that it is never created or destroyed; that is to say, that the number of nucleons must remain constant. Thus a proton can be created out of energy, but only together with an antiproton. The two cancel each other both mathematically and, when they come together, physically.

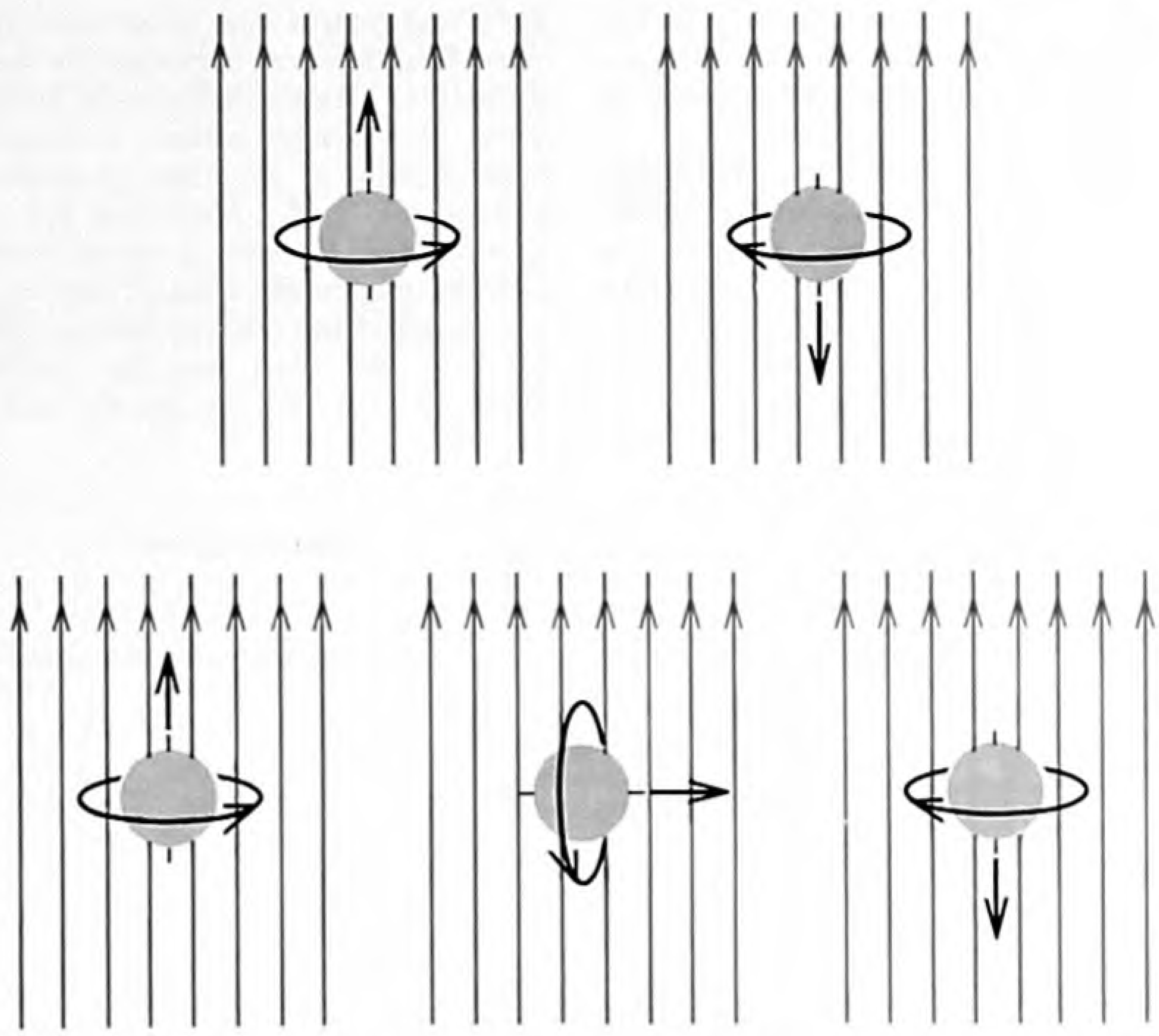
Particle Reactions

We should mention two other characteristics of particle reactions which appear to operate as general laws. First, the reactions are reversible. If one particle is observed to split into two others, we expect to find that the pair can also combine to form the original particle.

Second, the emission of a particle is related to the absorption of the corresponding antiparticle: if we know the "cross section" or probability for one, we can compute the probability for the other. For example, consider the beta-decay process [*see reactions 1-3 in the table on page 102*]. Since a neutron turns into a proton by emitting an electron and an antineutrino we expect to find the reverse reaction in which a proton absorbs an electron and an antineutrino, turning into a neutron [*reaction 2*]. Furthermore, since the absorption of an electron corresponds to the emission of an antielectron, or positron, we should also have the reaction in which a proton absorbs an antineutrino and emits a positron, again turning into a neutron [*reaction 3*]. This reaction is, in fact, the one by which the neutrino was finally detected experimentally.

Thus the rules provide a kind of algebra with which we can "solve" problems in particle physics. Let us see how they work for a sample problem—the decay of the pion. The neutral pion is found experimentally to decay into a pair of photons in a very short time: about 10^{-15} seconds. We want to show why it should do this [*reactions 4-8*]. We start out with the basic reaction involving pions, the Yukawa reaction in which a nucleon, let us say a proton, emits a virtual pion. We are interested in what the pion will do, so we should like to have an equation that starts out with a neutral pion followed by an arrow. By reversing the equation and transposing terms (changing particles into antiparticles) we arrive at the desired equation [*reaction 6*], which tells us that the pion turns (virtually) into a proton-antiproton pair. Such a pair annihilates, and can yield photons. Thus we have arrived at the desired result: a neutral pion decaying into two photons.

To someone seeing it for the first time, this chain of reasoning may seem just a trivial shuffling of symbols. But each shuffling summarizes a detailed (and often difficult) calculation of probabilities. Thus in the end we arrive at a reasonably exact prediction of what will happen, how long it should take, and so on. On the other hand, our knowledge of the situation is so incomplete that an apparently airtight chain of reasoning can also result in completely wrong answers.



SPINNING PARTICLE can take only certain fixed positions with respect to an outside magnetic field (*colored arrows*). Particles with spin 1/2 (*top*) can align their axes with or against the field. Particles with spin 1 (*bottom*) can go with, across or against the field.

The Muon

There is no better illustration of this than the decay of the charged pion. Using our basic reactions and rules we can "prove" that the positive pion should decay to a positron and a neutrino [*reactions 9-14*]. But if the reaction ever happens, it is so rare that it has never been observed. How does the positive pion actually decay? It yields a neutrino and a totally new particle: the muon!

Here we have nature at her most perverse. She has given us a particle for which there is no theoretical justification and no use whatever. The muon was the unwelcome baby on the doorstep, signifying the end of days of innocence. The situation was further complicated by an unfortunate historical accident which for a long time made it impossible even to classify the muon. The muon was detected before the pion, and everyone took it to be the meson Yukawa had predicted. For this role, however, its properties are all wrong. It notably does not interact strongly with nucleons and thus could not be the particle responsible for the nuclear-force field. Thus until the pion was discovered the muon made even less sense than it does today.

Now that we know what the muon is not, we can at least accept it for what it is. It comes with both positive and negative charge (the negative pion decays into a negative muon and antineutrino). It weighs about 207 electron masses and has a spin of 1/2 (*i.e.*, it is a fermion). It lives for about a millionth of a second and then decays into an electron, a neutrino and an antineutrino. The positive muon must of course yield a positive electron, or positron, and the negative muon a negative electron. Each muon is the other's antiparticle.

Although the muon does not come out of the dozen-particle theory—indeed, it demonstrates that such a theory is incomplete if not wrong—it can be connected to the other particles. To see the connection we must re-examine our fundamental processes—the Dirac or electromagnetic interaction, the Yukawa or nuclear interaction and the Fermi or beta-decay process. It turns out that these processes differ enormously in "strength." The Yukawa process is known as a strong interaction; it accounts for the great force that holds nucleons together in the nucleus. Electromagnetic forces are some 137 times weaker than nuclear forces. Another indication of strength is the probability that a process will occur in a given time; *i.e.*, its average rate. The strong interactions are as fast as anything can possibly be. The emission or absorption of a pion takes place in some 10^{-23} seconds, which is just about the time it

would take a light ray to cover a distance equal to the diameter of a nucleon. The electromagnetic process is of course 137 times slower.

What about the Fermi interaction? It is incomparably weaker than the others. The factor is about 10^{-14}—it is a hundred thousand billion times weaker than the strong interactions! Furthermore, all the processes involving neutrinos—beta-decay and the decay of pions and muons—are about equally weak. Thus the muon participates in one of the three fundamental types of interaction. (As a charged particle, of course, it participates in the electromagnetic interaction as well.) Also, since it is a light fermion, it seems to group itself naturally with the electron and neutrino.

A glance at the table of lifetimes on this page will show that the decays which we have said are equally weak have widely different times. But speed is supposed to be an indication of strength. The answer is that speed is not determined solely by strength. It also depends on the energy available to make the reaction go. In the case of the neutron, which takes an average of 18 minutes to decay, there is very little available energy; the difference between the mass of the decaying particle and the mass of its decay products is only slightly more than the mass of one electron. For pion and muon decays there is much more energy; they are correspondingly faster. If corrections are made in each case for the available energy, it turns out that a kind of "intrinsic" speed for all the weak processes is very close to 10^{-9} seconds, which is 10^{14} times slower than the strong interaction.

It is surely remarkable that all the weak processes have the same strength, and it is probably significant. Nature is trying hard to tell us something, but so far we have been unable to decipher the message.

Strange Particles

With the muon nature gently warned physicists that they had not yet divined her innermost secrets. Then around 1950 she rudely introduced a whole procession of new particles. They were utterly unexpected and had properties which could not be explained on the basis of previous theory.

The new arrivals showed up first in the showers of particles which occur when high-energy cosmic rays strike a lead plate inside a cloud chamber. Among the tracks of the showers were found some curious two-pronged or V-shaped patterns that could not be explained by any known particle process. Physicists were forced to conclude that some unknown neutral particle (which would leave no track in the cloud chamber) had decayed into two charged particles. The neutral particle presumably had been made in the lead plate. Once people started looking for the so-called V-particles, they turned out to be very common.

As V-events were collected and studied, it became clear that there were at least two new neutral particles. One which decays into a proton and a negative pion, was named the lambda; the other, which decays into a positive and a negative pion, was called the K.

When they had recovered from the shock, physicists began to try to fit the new particles somehow into the general scheme. From the pattern of its decay (into a fermion and a boson) the lambda

PARTICLE		SPIN	REST MASS (ELECTRON MASSES)	MEAN LIFE (SECONDS)	DECAY PRODUCTS
XI	Ξ^-	$\frac{1}{2}$	2585	10^{-10} TO 10^{-9}	$\Lambda^\circ + \pi^-$
	Ξ°	$\frac{1}{2}$	NOT YET FOUND		
SIGMA	Σ^+	$\frac{1}{2}$	2325	$.7 \times 10^{-10}$	$p + \pi^\circ$ $n + \pi^+$
	Σ^-	$\frac{1}{2}$	2341	1.5×10^{-10}	$n + \pi^-$
	Σ°	$\frac{1}{2}$	2324	NOT MEASURED	$\Lambda^\circ + \gamma$
LAMBDA	Λ°	$\frac{1}{2}$	2182	2.7×10^{-10}	$p + \pi^-$ $n + \pi^\circ$
PROTON	p	$\frac{1}{2}$	1836.1	STABLE	
NEUTRON	n	$\frac{1}{2}$	1838.6	ABOUT 1,000	$p + e^- + \bar{\nu}$
K MESON	K^+	0	966.5	1.2×10^{-8}	$\mu^+ + \nu$ $\pi^+ + \pi^\circ$ $\pi^+ + \pi^+ + \pi^-$ $\pi^+ + \pi^\circ + \pi^\circ$ $\mu^+ + \nu + \pi^\circ$ $e^+ + \nu + \pi^\circ$
	K^-	0	966.5	1.2×10^{-8}	$\mu^- + \bar{\nu}$ $\pi^- + \pi^\circ$ $\pi^- + \pi^- + \pi^+$ $\pi^- + \pi^\circ + \pi^\circ$ $\mu^- + \bar{\nu} + \pi^\circ$ $e^- + \bar{\nu} + \pi^\circ$
	K_1°	0	965	1×10^{-10}	$\pi^+ + \pi^-$ $\pi^\circ + \pi^\circ$
	K_2°	0	965	3×10^{-8} TO 10^{-6}	$\pi^+ + e^- + \bar{\nu}$ $\pi^- + e^+ + \nu$ $\pi^+ + \mu^- + \bar{\nu}$ $\pi^- + \mu^+ + \nu$ $\pi^+ + \pi^- + \pi^\circ$ $\pi^\circ + \pi^\circ + \pi^\circ$
PION	π^+	0	273.2	2.6×10^{-8}	$\mu^+ + \nu$
	π^-	0	273.2	2.6×10^{-8}	$\mu^- + \bar{\nu}$
	π°	0	264.2	10^{-16} TO 10^{-15}	$\gamma + \gamma$
MUON	μ^-	$\frac{1}{2}$	206.7	2.2×10^{-6}	$e^- + \nu + \bar{\nu}$
ELECTRON	e^-	$\frac{1}{2}$	1	STABLE	
NEUTRINO	ν	$\frac{1}{2}$	0	STABLE	
PHOTON	γ	1	0	STABLE	

PROPERTIES OF PARTICLES are being determined with increasing precision. This table presents the latest experimental values. Spins of the K mesons and the strange heavy particles are still doubtful. For the unstable particles which have more than one mode of decay the table lists all the sets of products now known. Others may still be discovered.

can be shown to be a fermion, presumably with spin 1/2. It is subject to the law of conservation of nucleons. Since one nucleon is produced in the decay, one must have been used in the formation process. For example, the lambda might be made in a collision between a proton and a negative pion—the reverse of the decay process. (Some other particle, such as a neutral pion, would also have to be made to carry off the excess energy.) The frequency with which lambdas appear shows that they are made by a strong process. The mass of the lambda turns out to be 2,181 electron masses.

The K particle has to be a boson because it decays into two pions, both bosons. Its spin must then be a whole number (most likely zero). It cannot be made out of a nucleon, because there are no nucleons in its decay products. It is, however, produced frequently, and thus by a strong process. It has a mass of 965.

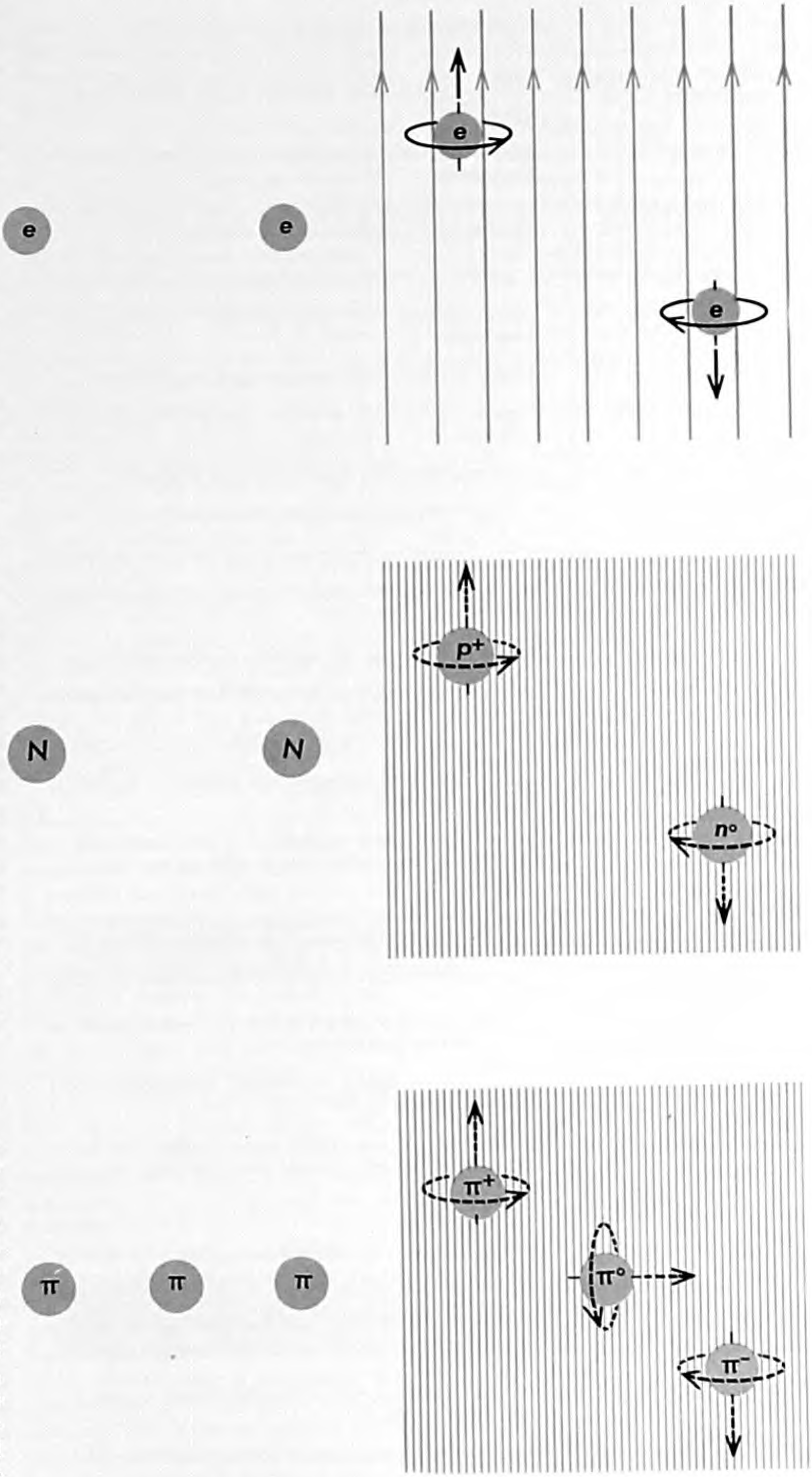
From these first considerations it is possible to classify the new particles to some extent. The lambda obviously belongs with the nucleons or heavy particles. It is made from a nucleon and, like them, it is a fermion. The K, on the other hand, is a boson; hence it is put in the meson group along with the pions.

The lambda and K particles were only the beginning. Soon other particles were identified. In the same category with the lambda are the sigma particles, charged and neutral, and the negative xi particle. Added to the neutral K were a pair of charged particles of about equal mass, called the positive K particle and the negative K particle.

Long Lifetimes

The very existence of all these distinct forms of matter is a difficult problem. But if we accept the fact that they exist, their behavior presents us with even deeper questions. The trouble arises with the decay of the new particles. Their lifetimes range from about 10^{-8} to 10^{-10} seconds, which is on the time scale of the weak interactions. But the particles are made, as we have seen, by strong interactions, the time scale of which is some 10^{-23} seconds. According to one of our most fundamental tenets—that of reversibility—a particle made in a strong interaction should also decay that way.

The new particles would seem to have ample opportunity to decay by strong processes. Consider, for example, the neutral lambda particle and let us play our equation-juggling game [*see reactions 15-19 on next page*]. By using two



DIFFERENT STATES of a particle can be recognized only through certain interactions. Two oppositely spinning electrons (*top*) would seem alike in the absence of an external magnetic field (*left*). When the field, indicated by colored arrows, is turned on, the electrons separate into different energy states (*right*). Similarly, in the absence of the electromagnetic interactions, all nucleons (*center*) are indistinguishable, as are all pions (*bottom*). When these interactions, indicated by closely spaced colored lines, are taken into account, isotopic "spin," which is represented by the broken arrows, separates the nucleons into protons and neutrons, and separates the pions into their three different charge types.

PARTICLE REACTIONS

1.	$n \rightarrow p + e^- + \bar{\nu}$	*Fermi process.* Neutron splits into proton, electron and antineutrino.
2.	$p + e^- + \bar{\nu} \rightarrow n$	*Reverse reaction.* Proton, electron and antineutrino combine into neutron.
3.	$p + \bar{\nu} \rightarrow e^+ + n$	*Transpose electron as antiparticle.* Proton interacts with antineutrino to yield positron and neutron.
4.	$p \xrightarrow{v} p + \pi^\circ$	*Yukawa process.* Proton splits (virtually) into proton and neutral pion.
5.	$p + \pi^\circ \xrightarrow{v} p$	*Reverse reaction.* Proton and pion combine into proton.
6.	$\pi^\circ \xrightarrow{v} p + \bar{p}$	*Transpose proton as an antiproton.* Pion splits into proton and antiproton.
7.	$p + \bar{p} \rightarrow \gamma + \gamma$	Proton and antiproton annihilate to yield photons.
8.	$\pi^\circ \rightarrow \gamma + \gamma$	*Net result.* Neutral pion decays to photons.
9.	$p \xrightarrow{v} n + \pi^+$	*Yukawa process.* Proton splits (virtually) into neutron and positive pion.
10.	$n + \pi^+ \rightarrow p$	*Reverse reaction.* Neutron and pion combine into proton.
11.	$\pi^+ \xrightarrow{v} p + \bar{n}$	*Transpose neutron as an antineutron.* Pion splits into proton and antineutron.
12.	$p + \bar{\nu} \rightarrow e^+ + n$	*Fermi process.* Proton interacts with antineutrino to yield positron and neutron.
13.	$p + \bar{n} \rightarrow e^+ + \nu$	*Transpose antineutrino and neutron as their antiparticles.* Proton interacts with antineutron to yield positron and neutrino.
14.	$\pi^+ \rightarrow e^+ + \nu$	*Net result.* Positive pion decays to positron and neutrino.
15.	$\pi^- + p \rightarrow \Lambda^\circ + \pi^\circ$	*Hypothetical process.* Negative pion and proton interact to yield lambda and neutral pion.
16.	$\Lambda^\circ + \pi^\circ \rightarrow \pi^- + p$	*Reverse reaction.* Neutral pion interacts with lambda to yield negative pion and proton.
17.	$\Lambda^\circ \xrightarrow{v} \pi^\circ + \pi^- + p$	*Transpose neutral pion (it is its own antiparticle).* Lambda decays (virtually) into a neutral pion, a negative pion and a proton.
18.	$p + \pi^\circ \rightarrow n$	*Yukawa process.* Proton absorbs neutral pion.
19.	$\Lambda^\circ \rightarrow p + \pi^-$	*Net result.* Lambda decays quickly to proton and negative pion.
20.	$\pi^- + p \rightarrow \Lambda^\circ + K^\circ$	*Hypothetical process.* Pion and proton interact to yield lambda and neutral K.
21.	$\Lambda^\circ + K^\circ \rightarrow \pi^- + p$	*Reverse reaction.* Lambda and K interact to yield pion and proton.
22.	$\Lambda^\circ \xrightarrow{v} \pi^- + p + \bar{K}^\circ$	*Transpose K as antiparticle.* Lambda decays (virtually) into pion, proton and anti-K.
23.	$p + \pi^- \rightarrow n$	*Yukawa process.* Proton absorbs pion, turning into neutron.
24.	$\Lambda^\circ \rightarrow n + \bar{K}^\circ$	*Net result.* Lambda decays into neutron and anti-K.

strong processes, one known and one hypothetical but plausible, we arrive at the statement that the lambda is converted to a proton and a negative pion, which is its actual mode of decay. There is plenty of energy available for the process: the lambda's mass is 74 units greater than that of the proton and pion, which gives an energy difference of 37 mev. Thus we have "proved" that the lambda must decay as fast as it is made. The same thing can be demonstrated for all the other new particles. The only trouble is that they live 100,000 billion times longer than they should! It was this enormous discrepancy between their expected and observed lifetimes that was chiefly responsible for the designation "strange" or "queer" particles.

Associated Production

After contemplating the situation for a couple of years a number of theoreticians, in particular A. Pais of the Institute for Advanced Study, were able to suggest a possible resolution of the paradox. Their idea was that strange particles are made only in groups of two or more at a time. The concept is now known as associated production. It implies that the strong interaction which manufactures a strange particle somehow works only on more than one at a time. The trick here is that a strong process of this kind would not be reversible because of lack of energy.

For example, suppose that a lambda and a K were made in the collision of a negative pion with a proton. Now we apply our reaction rules to this process in order to predict the fate of the lambda [*reactions 20-24*]. We arrive at the conclusion that it "decays" into a proton and an anti-K. But of course this is impossible, because the two daughter particles have a combined mass greater than that of the parent. A thorough analysis shows that every possible case of associated production leads to a similar result for the separate decay of any one of the strange particles that are made. The possible avenues of decay always turn out to require too much energy. Thus, by moving away from each other immediately after they are created, the strange particles are saved

PARTICLE EQUATIONS summarize certain detailed chains of reasoning discussed in the text. Some of the chains of reasoning are found to be faulty, as indicated by the fact that the reactions they predict do *not* take place. These have been shown in gray.

from death by strong interaction, and they live until the much less probable weak processes catch up with them.

At first there was little or no experimental evidence for associated production. However, when the Cosmotron at the Brookhaven National Laboratory began to make strange particles on order, it appeared that the rule was indeed obeyed. In fact, the very first reaction to be discovered was the pion-proton collision suggested above, leading to the associated production of the lambda and the neutral K particles.

Now the question arose: What does associated production mean? Can it be related to any other principles? It tells us that strong reactions involving single strange particles are forbidden. When nature rules out an event, her legislation often takes the form of a conservation law. Such-and-such cannot happen because something must be conserved. To take a simple example, we never see a particle decay into products whose total mass is greater than its own. The conservation of energy forbids it.

Once the rule of associated production had been found, it was natural to ask whether there might not be a conservation law behind it. If the law could be discovered, we might find out much more about strange particles. Associated production says they must be made more than one at a time. But are all combinations possible or are some ruled out? The conservation law should tell.

Isotopic Spin

It appears that the law has been discovered, and we do know some of the rules by which strange particles are made. In order to see what the principle and the rules are, we must go back to the older particles and to a concept which we have not mentioned until now. This is the notion of "isotopic spin."

First let us have another look at ordinary spin. Imagine that we have a pair of isolated electrons which we can actually see as small specks. So far as we can tell, they are identical. We believe they are spinning, but they are so small that we cannot detect their motion. Now we put them in a magnetic field. Obeying the laws of quantum mechanics, their spins line up with or against the field. Suppose one goes with and the other against. Then the two particles have different energies, and we can distinguish one from the other. This imaginary experiment underlines the fact that the electron is a "doublet" so far as its magnetism is concerned. It may be in one of two possible energy states. But without an external magnetic field there is no way to tell the states apart; they "degenerate" into indistinguishability.

Now in the early days of modern nuclear physics—soon after the discovery of the neutron—a situation arose that was reminiscent of this magnetic "degeneracy." Experiments on the deflection of moving protons and neutrons by other protons and neutrons disclosed the surprising fact that the nuclear force, or strong interaction, between nucleons is the same regardless of the type of particle involved. The forces between two protons, two neutrons, or a proton and a neutron are all equal. This phenomenon, called charge independence, means that so far as strong interactions are concerned the neutron and proton look like the same particle. They can be distinguished only by their electromagnetic interaction. Suppose electromagnetism could be "turned off" like a magnetic field in the laboratory. Then the proton and neutron would also degenerate into indistinguishability. Hence the nucleon can be thought of as a "charge doublet," with one state representing the proton and the other the neutron.

This idea occurred to Heisenberg, who proceeded to express it mathematically. He constructed a mathematical description of the nucleon which included a variable that could take on just two values. One value thus represents the proton and the other the neutron. The mathematics is very much like that used by Pauli to describe the spin of an electron. Thus Heisenberg called his quantity isotopic spin. "Isotopic" refers to the fact that in a sense the proton and neutron are isotopes: they have nearly the same mass but different charge. "Spin" is purely an analogy, and a somewhat misleading one at that. It simply reflects a similarity to the mathematical term for real spin.

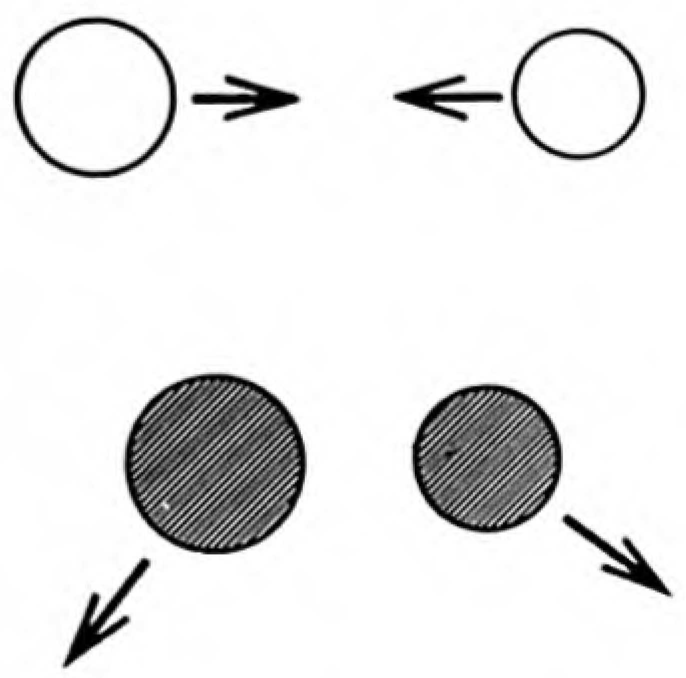
Isotopic spin, then, is a mathematical device which distinguishes the proton and neutron; physically they are distinguished by their different couplings to the electromagnetic field. The analogy to real electron spin is very close: the isotopic spin of the nucleon is also 1/2. Like real spin, it has components +1/2 and −1/2 with respect to a given, or reference, direction. In quantum electrodynamics it is customary to place a particle in a system of coordinates, the "Z" axis of which is parallel to the surrounding magnetic field. Hence the reference direction for isotopic spin is also considered to go along the Z axis, and the components are denoted by I_z. The convention has been adopted that I_z of +1/2 represents the proton and I_z of −1/2 the neutron. In Heisenberg's mathematical theory charge independence becomes a conservation law. When nucleons interact, the total isotopic spin is conserved. This statement can be shown to mean the force is equal between a proton and a proton, a neutron and a neutron, and a proton and a neutron. So far, it should be emphasized, the idea of isotopic spin is a mere formality. It adds nothing to the notion of charge independence, and is simply another convenient way of expressing it mathematically.

Charge Multiplets

When Yukawa explained nuclear forces in terms of pion emission and absorption, isotopic spin took on a somewhat broader significance. The British physicist Nicholas Kemmer, now at Edinburgh University, realized that the concept must also apply to pions. His reasoning was as follows: Nuclear forces, involving the virtual exchange of pions, are charge-independent. Therefore pions must be charge-independent, and the isotopic spin concept should apply.

Now we recall that the pion has three possible charges: plus, minus and neutral. Thus it constitutes a charge "triplet" which, if charge could be turned off, would also degenerate into indistinguishability. In the case of real spin, a triplet means that the particle has a spin of 1, since it may then assume three different directions with respect to the field. Its Z components are +1, 0 and −1. So we assign an isotopic spin of 1 to the pion, and we say that its components with respect to the reference direction (its I_z's) are +1, 0 and −1.

The grouping of particles into charge doublets or triplets (collectively called multiplets) provides a convenient short-



ASSOCIATED PRODUCTION is suggested in this diagram. Two normal particles collide (*top*), making a pair of strange particles (*bottom*) which immediately separate.

hand way of identifying them. If we say that the pion is a triplet with its center of charge at zero [*see chart on opposite page*], this tells us at once that the charges are +1, 0 and –1, the isotopic spin is 1, and the I_z's are +1, 0 and –1. Similarly, to say that the nucleon is a doublet with center of charge at +1/2 means that the charges are 0 and +1, the isotopic spin 1/2 and the I_z's +1/2 and –1/2. As can be seen in the chart, the antinucleon is another doublet centered at –1/2. Its isotopic spin is 1/2 and its I_z is –1/2 and +1/2.

Note that the isotopic spin and charge multiplet concepts give us another difference between nucleons and pions. Nucleons are a doublet centered at +1/2, while the pions are a triplet centered at 0.

Now let us turn to the strange particles. An obvious question is whether their interactions are also charge-independent and conserve isotopic spin. Are positive sigma particles, for example, just like negative and neutral ones except for their electric properties? There is no direct experimental evidence on the point, but it seemed reasonable to suppose that charge independence would apply to the strong interactions of the new particles as it had been found to apply to the strong Yukawa coupling. This would mean that the strange particles are charge multiplets. If so, it was generally supposed that they would follow the same classification as the nucleon and pion. That is, the heavy strange particles seem to be related to nucleons; they are made out of nucleons and decay back to nucleons. Therefore the heavy particles were generally thought to be doublets, having an isotopic spin of 1/2 and a charge center at plus or minus 1/2. The K particles, on the other hand, apparently belong with the pion, so it was supposed they would fall into a triplet, with an isotopic spin of 1 and charge center at 0.

About five years ago one of the authors of this article (Gell-Mann) and the Japanese physicist Kazuhiko Nishijima independently conceived the idea that the strange particles might not follow this arrangement. Furthermore, the departure from the expected arrangement might account for their strange behavior. In the case of the present author it was a matter of discovery by slip of the tongue. Discussing the heavy strange particles one day, he spoke of them as having an isotopic spin of 1, but then quickly corrected himself, saying "I mean a half, of course."

The more he thought about the "mistake" later on, the more he began to wonder whether it really was one. How do we know that heavy particles are doublets with isotopic spin 1/2? To be sure, the particles seemed to be related to the nucleon—and for the sake of order and simplicity one certainly hoped that they were related. But if they were members of that family, they were strange members. Perhaps it was precisely in their isotopic spin that their strangeness lay. Suppose the heavy particles, instead of being doublets of isotopic spin 1/2, like the nucleon, were triplets of isotopic spin +1 or even singlets of isotopic spin 0. (A particle with zero isotopic spin has only one possible state and is thus a singlet.) Suppose the K particles, instead of being triplets like the pion, were doublets? (At the time the table of strange particles was just being filled in by experiment. It was not even known, for example, whether there were charged K or lambda particles.)

After toying with the idea for a while, the author began to see that it might contain in it just the conservation law that was needed to explain associated production and the strangely long lifetimes of strange particles. In a moment we shall try to show roughly how this comes about. First let us pursue the idea a little further.

Displaced Multiplets

Recall that a simple way of describing a group of particles is to indicate its center of charge and whether it is a doublet or triplet. The nucleon is a doublet with center at +1/2; the pion is a triplet with center at 0, and so on. Now suppose that among the heavy particles there is a singlet at charge 0 [*see chart on opposite page*]. Could this, by any chance, be the neutral lambda? If it is, note that the center of this "multiplet" (a singlet is a multiplet with one member) is at 0—one-half charge unit less than the center of the nucleon doublet. The original expectation was that all heavy particles would have their multiplet centers at +1/2. Therefore the lambda is "displaced" by –1/2 charge unit. Perhaps this displacement is the essential physical characteristic of the particle which accounts for its "strangeness." Let us assume that it is; in fact, let us invent a new physical quantity and call it strangeness. For reasons of mathematical convenience we make the strangeness equal to twice the displacement. Thus the strangeness of our putative lambda particle is twice the displacement of –1/2, or –1. (The strangeness of the nucleon is of course 0. Its charge center sets the reference point from which the displacements of other heavy particles are measured.)

Next we observe that, in our system of classification by multiplets, the antinucleons form a doublet which is an image of the nucleon doublet, mirrored on the zero-charge line [*see chart*]. Thus the other heavy particles should also have antiparticles in corresponding multiplets. We accordingly place in our table an antilambda, which is also at 0. Its displacement is +1/2 (from the "normal" charge center of the antinucleon). Hence its strangeness is +1.

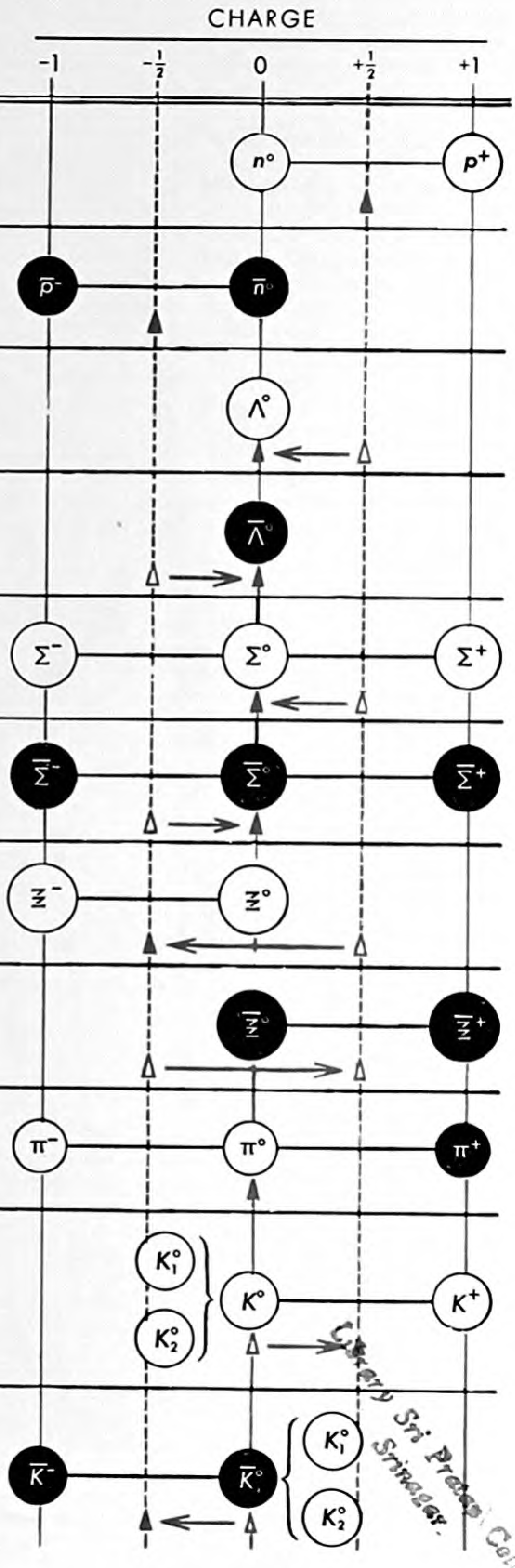
Now we can try something else—say a triplet centered at 0. Its strangeness would be –1. If there is such a triplet, there should be three strange particles, positive, negative and neutral, all with approximately the same mass. At the time the strangeness theory was conceived, no such triplet was known. Now it has apparently been found in the sigma particle (Σ^+, Σ°, Σ^-). Again we expect a corresponding multiplet of antiparticles.

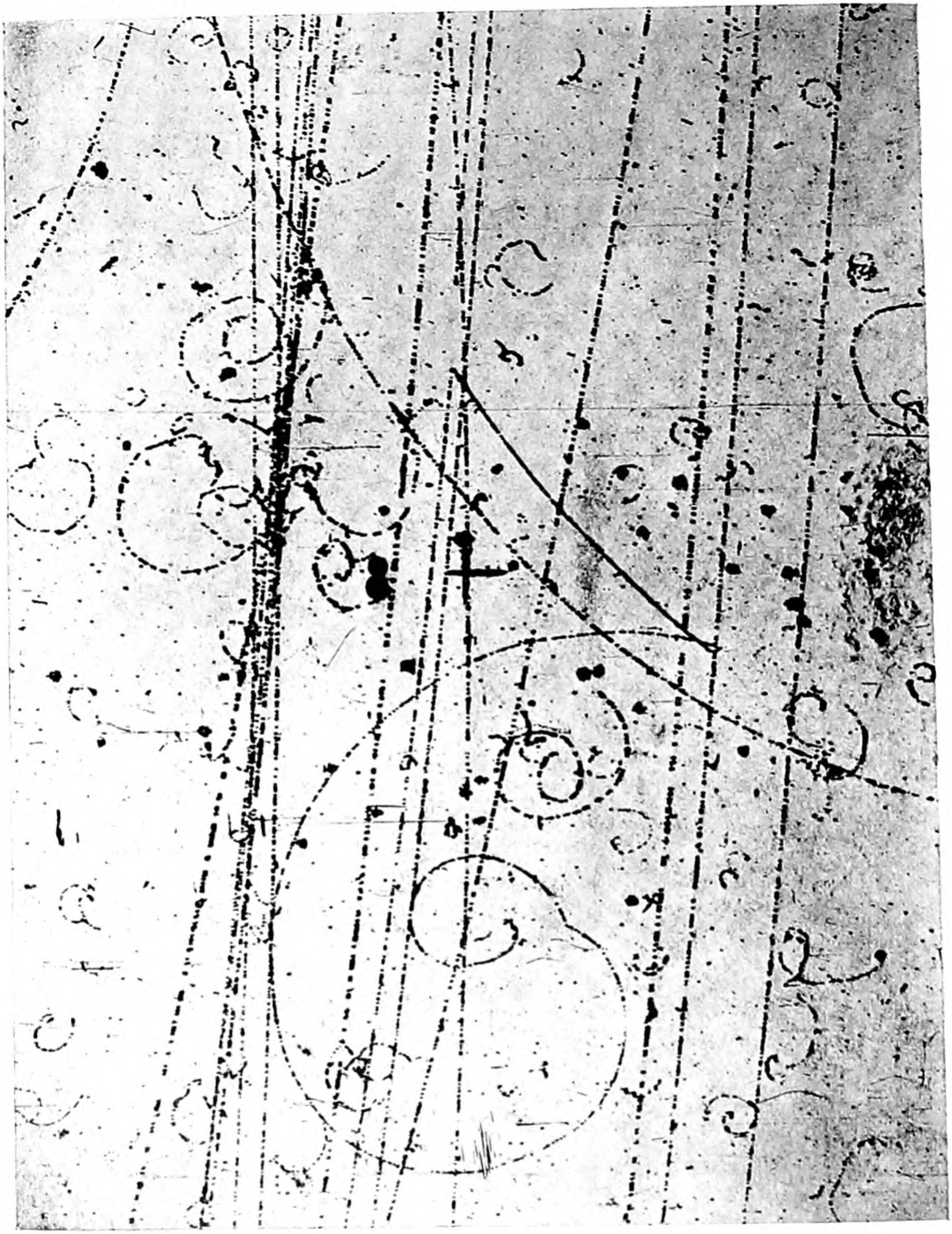
Still another possibility is a heavy-particle doublet displaced a full charge unit, from +1/2 to –1/2, *i.e.*, with strangeness –2. This would mean a pair of particles with charge –1 and 0. We now believe that the negative member of the pair is the xi particle (Ξ^-). The neutral member (Ξ°) has not yet been detected, but the general success of the strangeness theory gives us considerable confidence that it will turn up.

The K particles may fall into doublets like the nucleon and antinucleon. This would mean that the K^+ and K° constitute one doublet with its charge center at +1/2. Since these particles are grouped with the pion, whose "natural" charge center is at zero, their displacement is +1/2 and their strangeness +1. The K^- then is part of a doublet together with a second neutral K, the anti-K°, which is the antiparticle of the first. The

STRANGENESS is illustrated in tabular form. Particles (*white circles*) and antiparticles (*black circles*) are grouped in multiplets with their charges indicated by the colored vertical lines. The solid colored carets mark the charge center of each multiplet; open carets mark the "expected" location of charge centers (1/2 for heavy particles, –1/2 for heavy antiparticles and 0 for mesons). Horizontal colored arrows show the displacement of each center from the expected position. The strangeness equals twice the value of this displacement.

PARTICLE	ISOTOPIC SPIN	STRANGENESS
NUCLEON	$\frac{1}{2}$	0
ANTI-NUCLEON	$\frac{1}{2}$	0
LAMBDA	0	-1
ANTI-LAMBDA	0	+1
SIGMA	1	-1
ANTI-SIGMA	1	+1
XI	$\frac{1}{2}$	-2
ANTI-XI	$\frac{1}{2}$	+2
PION	1	0
K	$\frac{1}{2}$	+1
ANTI-K	$\frac{1}{2}$	-1





K-particle formation is shown as a series of bubbles produced by charged particles in a chamber of liquid propane. The positive pion decays further to a muon and a neutrino, which leaves no bubble track. Finally the muon decays to a positron and two invisible neutrinos. A neutral lambda was presumably made together with the K, but if so it left the chamber without decaying to charged particles and made no track. The incoming pion was produced in the Cosmotron at Brookhaven National Laboratory.

center of this doublet is –1/2, its displacement is –1/2 and its strangeness is –1. The pion plays the same role in the meson group as the nucleon does in the group of the heavy particles. That is, its charge center provides the reference point for measuring strangeness, so its own strangeness is 0.

Conservation of Strangeness

Now we have assigned a strangeness to all the strongly coupled particles. What is the point of this exercise? Simply this. It is possible to prove from the principle of charge independence that strangeness must be conserved in the strong and electromagnetic interactions. That is to say, in any reaction of these two types the total strangeness of the particles entering the reaction must equal the total strangeness of the products. We will show that this conservation law accounts for the observed behavior of strange particles.

To begin with, it obviously "explains" associated production. Strange particles are made in collisions between ordinary particles. The strangeness of the latter is 0. Therefore the total strangeness of the products must be 0. This means that at least two must be made at a time so that their individual strangenesses offset each other. Consider the case we have already mentioned: the production of a lambda and a neutral K from the collision of a pion and a proton. The lambda has a strangeness –1 and the K° has strangeness +1: total strangeness, 0.

As we have already seen, associated production explains why strange particles do not decay by strong interaction. But they must also be immune to the electromagnetic process, since their lifetimes are on the time scale of the weak interactions. The law of conservation of strangeness shows us how the particles avoid decay by electromagnetism as well as by strong interactions.

We can indicate only crudely how the law does so. It can be shown that the conservation of strangeness is mathematically equivalent to the conservation of the Z component of isotopic spin: I_z. The latter quantity is essentially a measure of charge: in any multiplet, the greater the I_z, the greater the charge. (For example, in the nucleon doublet I_z's of –1/2 and +1/2 correspond to charges of 0 and +1; in the pion triplet I_z's of –1, 0 and +1 correspond to charges of –1, 0 and +1, etc.) Electromagnetic interactions are thought to depend only on charge. By a general rule of quantum mechanics this means that they should conserve I_z (which measures charge). But to say they conserve I_z is the same as saying they conserve strangeness. Hence an isolated particle whose strangeness is not 0 cannot decay into particles with zero strangeness by an electromagnetic process.

Eventually, of course, the strange particles do decay into ordinary ones, and on a time scale about the same as that for the weak interactions. So it seems that the strange-particle disintegrations are members of this great class of processes. Weak interactions, therefore, do not conserve strangeness. It has recently been discovered that they also violate another conservation law, the conservation of parity, which has to do with symmetry in nature between right and left [see "The Overthrow of Parity," by Philip Morrison; SCIENTIFIC AMERICAN Offprint 231]. Whether there is any connection between the two laws and their violation is not known. In any case, it is clear nature has been concealing some of her most important secrets in the weak processes, and that one of the major jobs facing physics today is to discover the laws which govern these processes.

Selecting Particles

When we began the search for a conservation law to account for associated production, it was with the hope that it would tell us still more about the birth and death of strange particles. And so it does. For instance, the rule of associated production would permit a reaction in which two neutrons collide to form a pair of lambdas.

In fact, this was considered one of the more likely reactions. But the reaction has never been observed, and the conservation of strangeness tells us that, practically speaking, it never will be observed. The neutron's strangeness is 0 and the lambda's is –1. Thus the neutron collision, if it yields a lambda, must also yield another particle with a strangeness of +1, such as a neutral K: for example ($n+n\rightarrow\Lambda^\circ+n+K^\circ$).

Once again let us consider the case of the sigma and K particles. The sigma is a triplet with strangeness –1. The K particles are a pair of doublets; the pair including the K^+ has a strangeness of +1, while the pair including the K^- has strangeness –1. Hence it is possible to make a Σ^- (strangeness –1) and a K^+ (strangeness +1) together, but not a Σ^+ and a K^-, both of which have strangeness –1. The first reaction has been discovered in pion collisions with protons ($\pi^-+p\rightarrow\Sigma^-+K^+$). Aside from the conservation of strangeness, there seems no reason why the reaction should not also go the other way ($\pi^-+p\rightarrow\Sigma^++K^-$). But this has not been found to happen, and the strangeness principle tells us why it has not.

As a still further example of the power of the strangeness rule, we may examine the decay of the neutral sigma. The sigma triplet has the same strangeness (–1) as the lambda. The sigma is also some 150 electron masses heavier. Therefore it should be possible, both from the standpoint of available energy and of conservation of strangeness, for the sigma to decay to the lambda. That is to say, the sigma should not have to wait for the weak processes to end its life. However, in the case of a charged sigma, there would have to be some other charged particle in the decay products to conserve charge. It might be a pion ($\Sigma^+\rightarrow\Lambda^\circ+\pi^+$). The pion's strangeness is 0, so the strangeness accounts are in balance. But the pion's mass is 270, or about 120 more than is available energetically. The neutral sigma, on the other hand, does not need to produce any other charged particles. Its excess energy can be carried away by photons ($\Sigma^\circ\rightarrow\Lambda^\circ+\gamma$). This reaction has in fact been observed. It is an electromagnetic process (since it involves photons) and is therefore only a little slower than the strong interactions themselves.

Thus strangeness gives us rules for selecting the possible strange particles and their possible decays. As a matter of fact, a few particles were predicted by the strangeness table before they were actually found. The only one still missing is the neutral xi.

The Neutral K

Before leaving this "periodic table" of strange particles, one final comment is in order. It will be noticed that the K° and its antiparticle are also listed as a different pair of particles called the K°_1 and K°_2. One of the most striking successes of the strangeness theory was the prediction of this situation. The reasoning which led to the prediction is too complicated to set forth here, but it indicates a remarkable shuffling process on the part of nature. The K° and anti-K° are made in different processes. Once made, each of them can decay in two different ways, one of which takes a little longer than the other. Quantum theory shows that only half of each type of particle can follow either mode of decay. Thus we have two different manufacturing processes and two different decay processes, with a reshuffling in between. Nature segregates the neutral K particles

on one basis in their manufacture and on another in their decay. She makes K°'s and anti-K°'s. After they are made, half of each of these particles "become" K°_1's and half become K°_2's. This is demonstrated by the way they decay.

In the strangeness theory, then, we have a means of classifying strange particles. The theory is consistent with the fundamental idea of four groups of particles and three types of reaction. Thus we still have only the heavy particles (some of them strange), the mesons (some of them strange), light particles and the photon. And the couplings between them are strong, electromagnetic or weak.

At present our level of understanding is about that of Mendeleyev, who discovered only that certain regularities in the properties of the elements existed. What we aim for is the kind of understanding achieved by Pauli, whose exclusion principle showed *why* these regularities were there, and by the inventors of quantum mechanics, who made possible exact and detailed predictions about atomic systems.

We should like to know the laws of motion of the particles; to predict, among other things, how they will interact when they collide and how these interactions will deflect one particle when it collides with another. As this article is written a number of physicists are hard at work on theories which they hope may supply the laws. Time will be the judge.

On a still more fundamental level there are questions to which the answers seem as yet much more remote. Are all the particles we have mentioned really elementary, or are some of them just compounds of other particles? If so, which are elementary and which are not? Why has nature chosen to use this particular set of particles to build the material world? Why are the charges of elementary particles limited to the values +1, –1 and 0? These and many other such puzzles seem to lie entirely beyond the power of our present theories. Shall we ever know the answers? Every physicist has an abiding faith that we shall. But it will probably require some wholly new ideas. For one thing, many theoreticians believe that the present concepts may be entirely inapplicable at extremely short distances —of the order of the dimensions of the particles. In fact, it is suspected that here these concepts become self-contradictory.

It is likely to be quite a while before the particle physicist finds himself out of a job.

The Authors

MURRAY GELL-MANN and E. P. ROSENBAUM are, respectively, a theoretical physicist and a member of the Board of Editors of SCIENTIFIC AMERICAN. Gell-Mann was born in New York City, did undergraduate work at Yale University and took his doctorate at the Massachusetts Institute of Technology in 1951. Later that year he attended the Institute for Advanced Study in Princeton, after which he went to the University of Chicago's Institute for Nuclear Studies. Gell-Mann's interest in mathematics and physics was first stimulated by his father, whose hobby is mathematics. At M.I.T. he studied with Victor F. Weisskopf; at Chicago he had an opportunity to observe Enrico Fermi at work. Of Fermi, Gell-Mann writes: "His great genius was not something he could share with others, but he could communicate something of his style." Gell-Mann likes to travel; he worked out many of his ideas on "strange" elementary particles during trips through the U.S. and Europe. He is now professor of physics at the California Institute of Technology.

Bibliography

CHARGE INDEPENDENCE THEORY OF V PARTICLES. Kazuhiko Nishijima in *Progress of Theoretical Physics*, Vol. 13, No. 3, pages 285-304; March, 1955.

ELEMENTARY PARTICLES. Enrico Fermi. Yale University Press, 1951.

THE INTERPRETATION OF THE NEW PARTICLES AS DISPLACED CHARGE MULTIPLETS. M. Gell-Mann in *Nuovo Cimento*, Supplemento al Vol. 4, Series 10, No. 2, 2° Semestre, pages 848-866; 1956.

SCIENTIFIC AMERICAN February 1955, Vol. 192, No. 2, pp. 46-50 OFFPRINT 214

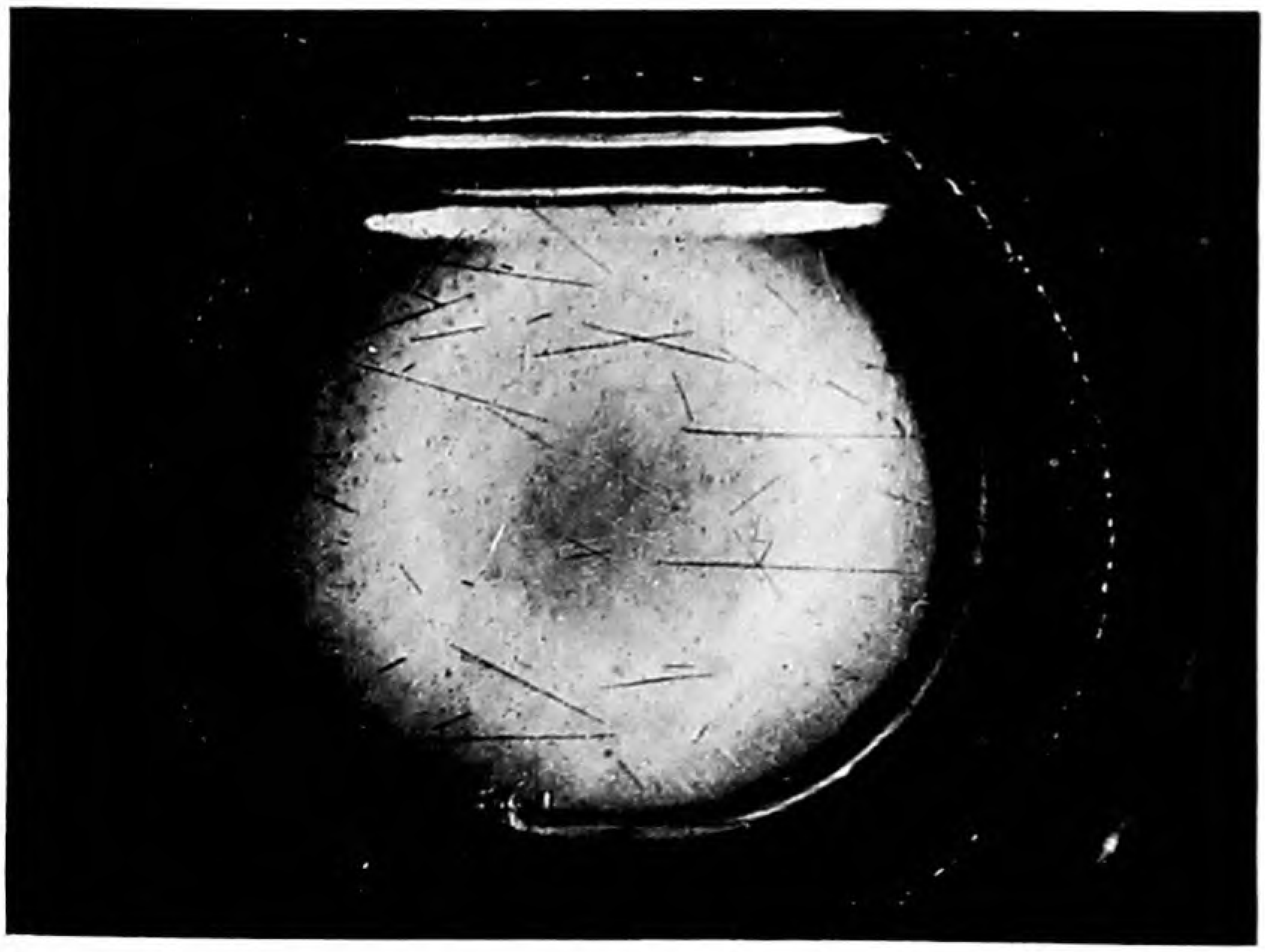
THE BUBBLE CHAMBER

by Donald A. Glaser

In their tense and ticklish state, superheated liquids will erupt into violent boiling when triggered by a nuclear particle. The first bubbles thus formed can be photographed to record the particle's adventures.

In their exploration of the submicroscopic world of atomic nuclei, physicists are like men groping in a dark cave with a flashlight that goes on for only an instant and each time lights only a tiny corner of the cave. Occasionally the flash catches some activity or event—either a familiar particle behaving in a familiar way or some strange new particle whose behavior is altogether baffling. From these scanty glimpses nuclear physicists are attempting to identify the particles and the forces at play in the dark, violent world of the nucleus of the atom. It would help if they had a better flashlight.

Let us look for a moment at the events they are trying to observe and at the observing devices that have been available up to now. Physicists are probing the nucleus by bombarding it with particles, preferably particles with enough energy to break up the nucleus into its constituent parts. The projectiles may come from cosmic rays, particle accelerators in the laboratory, nuclear reactors or other sources. When a high-energy particle hits the target nucleus, any of a number of different things may happen: it may ricochet off and leave the struck nucleus unchanged; it may break the nucleus into fragments; it may give birth to entirely new particles, or it may be absorbed. These events generally take place in something less than a millionth of a second. From that very brief glimpse the physicist seeks to determine the energy, the electric charge, the size or mass and the forces of interaction among the particles involved.



BUBBLE TRACKS in liquid hydrogen were produced by Luis W. Alvarez and his group at the University of California. Their 2.5-inch chamber, operated at about 400 degrees below zero Fahrenheit, was exposed to neutrons. Tracks show paths of recoiling protons.

He has had two ways of seeing and measuring these happenings. The first is the Wilson cloud chamber. In a chamber supersaturated with a vapor, a flying charged particle leaves a visible trail of liquid droplets, which condense on the ions the particle has produced by hitting vapor and gas atoms in its path. The density of droplets along the particle's track indicates its velocity and charge; its response to a magnetic field shows its momentum and whether its charge is positive or negative; its occasional disruption of the nucleus of an atom in its path, or its rebound from such a nucleus, tells something about the forces in the nucleus. Sometimes the particle breaks down ("decays") into lesser particles which make divergent tracks. But these interesting events occur only rarely in a vapor-filled chamber, because collisions in the gas are infrequent. To improve the chances of the particle hitting a nucleus, one can put a series of lead plates in the chamber or increase the pressure (*i.e.*, density) of its gas. Nonetheless, the cloud chamber has important limitations. The density of the track is very difficult to measure accurately. Particle tracks are lost from sight in the lead plates, so that the length of the track cannot be precisely calculated and details of collisions are hidden. After one particle has passed through, it takes a comparatively long time to clear the chamber for the next event, especially when the chamber is pressurized.

The second device for recording nu-

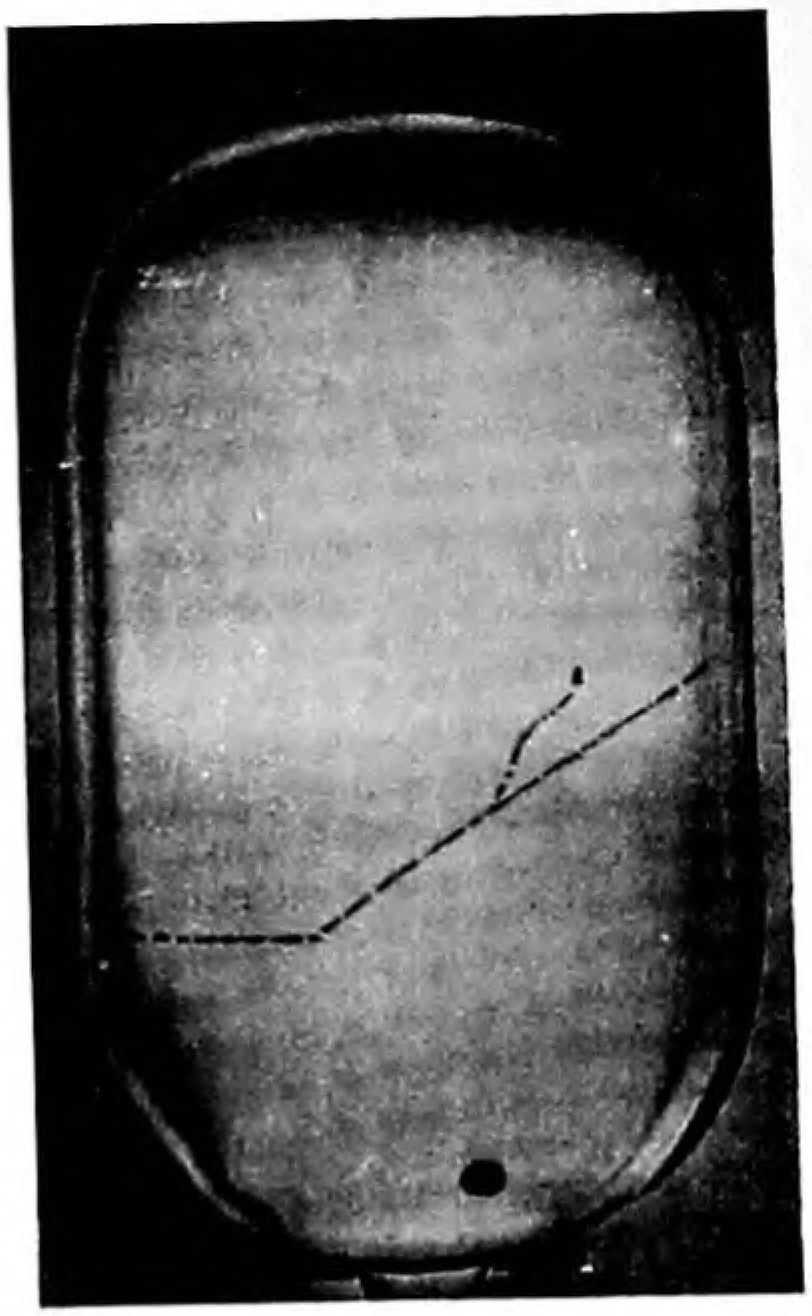
clear events is the photographic emulsion. It goes to the other extreme: it traps particles in a solid instead of in a vapor. In passing through the emulsion a particle leaves a trail of "exposed" silver grains which shows up on development. Nuclear physicists use special thick emulsions, stacked up in a number of layers to lengthen the tracks of particles passing through. The nuclear emulsion has several advantages: It is simple, compact and can be carried anywhere, even into the upper atmosphere by balloons and rockets. A particle charging into the dense emulsion has a high probability of colliding with nuclei; hence there is a good chance that the emulsion will show interesting events, including scattering, disintegrations and the formation of new particles. However, the emulsion also has its drawbacks. Its very density makes collisions so frequent and the particle path so crooked that the effect of a magnetic field cannot be measured. And from the moment it is manufactured an emulsion begins to collect random particle tracks from cosmic rays and terrestrial radioactivity; by the time the film is developed it is difficult to disentangle separate events and impossible to tell when the various tracks were made. Moreover, it takes tedious scanning under a high-power microscope to find the tracks at all.

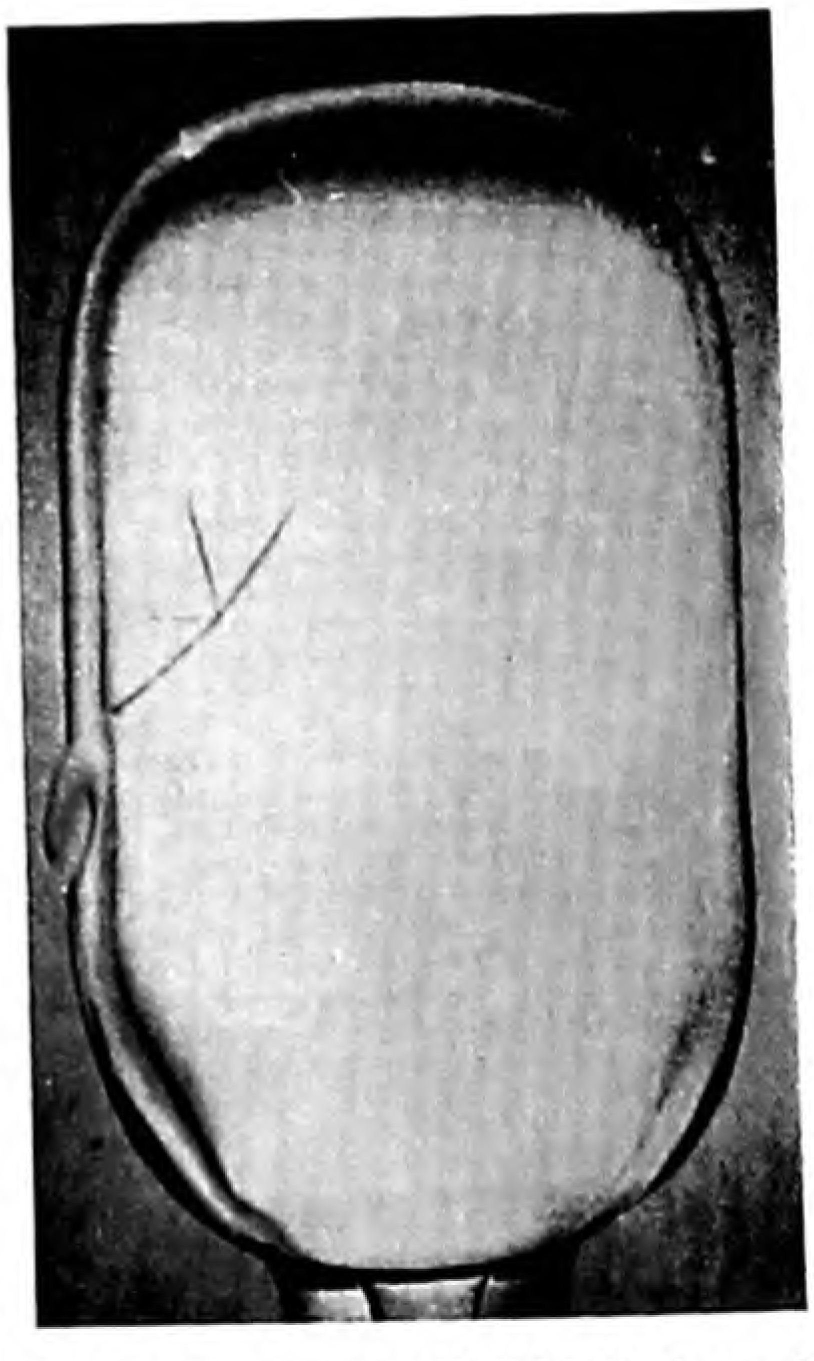
Could some compromise be found which would eliminate the defects and combine the respective virtues of the cloud chamber and the emulsion? It would be handy to find a particle-trapping medium which would be dense enough to afford the frequent collisions and precise tracks of the emulsion, and flexible enough to be amenable to magnetic fields and to give the good-sized, single picture of the cloud chamber, with a quick recovery after each exposure. In May, 1952, I began to try a new approach to the problem, and I soon decided to explore the possibility of a liquid medium.

What kind of reversible process in a liquid could show the path of a flying particle and quickly erase the track after its passage? It would have to be a process that magnified the tiny effect of the atomic particle itself, as the condensation of droplets in supersaturated vapor magnifies the ionization produced by a particle in a cloud chamber. It occurred to me that a superheated liquid, like a supersaturated vapor, might provide the desired unstable equilibrium that could be triggered by a small stimulus to yield a large effect. Physical chemists have long known that in a clean, smooth-walled vessel a very pure liquid may be heated above its usual boiling point without boiling. When the superheated liquid does begin to boil, it erupts with considerable violence, sometimes smashing the vessel. In chemical processes subject to this hazard bits of broken glass or other "boiling stones" are often thrown in to provide triggering points for boiling and thus prevent superheating. I wondered whether a flying particle might, under suitable conditions, trigger the formation of the microscopic bubbles that start the boiling process. If so, it might make a visible track in a superheated liquid.

Before undertaking any experiments, it seemed worth while to consider the question theoretically to see whether the idea could possibly work and what liquids would be most promising to try. I started with the supposition that a charged particle passing through the liquid might produce small clusters of ions of the same sign (negative or positive). Pushed apart by their mutual repulsion, the ions might leave tiny cavities, or bubbles, in the liquid. If such a mechanism actually operated, the specifications for a liquid that could record the tracks of particles were clear: the liquid should be nonconducting, so that the ions would retain their charges; it should have low surface tension, so that the force tending to collapse a cavity would be weak, and it should have high vapor pressure, which would tend to enlarge each cavity formed in the liquid. Worked out in detail, the theory gave the rather sharp prediction that ether (the ordinary anesthetic), when superheated to about 285 degrees Fahrenheit, should work as a track-forming liquid.

I made the first test very simple, because the whole idea seemed a long shot. I sealed some pure ether in an apparatus consisting of two narrow glass bulbs (about four inches long and a tenth of an inch in inside diameter) connected by a capillary tube. The vessel was not completely filled with liquid: there was a little space left for vapor from the liquid ether. Now one of the bulbs was immersed in a hot bath at 285 degrees and the other heated to 320 degrees. As a result, the higher vapor pressure in the hotter bulb forced liquid into the other one so that the latter was filled with liquid. Since the heating raised the pressure within the vessel to considerably more than that of the atmosphere, the liquid did not boil. Then the hotter bulb was removed from its





PENTANE-FILLED CHAMBER designed by R. H. Hildebrand and his colleagues at the University of Chicago demonstrates that different types of particles produced recognizably different tracks. In the top photograph a pi meson from a synchrocyclotron enters at the left, undergoes an elastic scattering, then ejects an electron (*wavy track*) from a pentane molecule. In the bottom photograph an incoming pi meson explodes a carbon nucleus and three slow, heavy charged particles are emitted. The density of bubbles along a track, which may be as high as 250 to the inch, is an index of the ionizing power of the passing particle.

hot bath. Its vapor quickly cooled and the pressure dropped to one atmosphere. The superheated liquid in the 285-degree bulb remained quiet for as long as several minutes. But when a bit of radioactive cobalt was brought near, it erupted instantly into violent boiling!

Further experiments with other liquids and other sources of radiation confirmed this discovery: ionizing radiation could indeed trigger boiling in a superheated liquid. We cannot say this proves the theory that electrostatic force forms the bubbles; it may be that the dash of a particle into the liquid simply adds a little triggering heat, like the plunge of a hot needle. But in any case the triggering action was a fact.

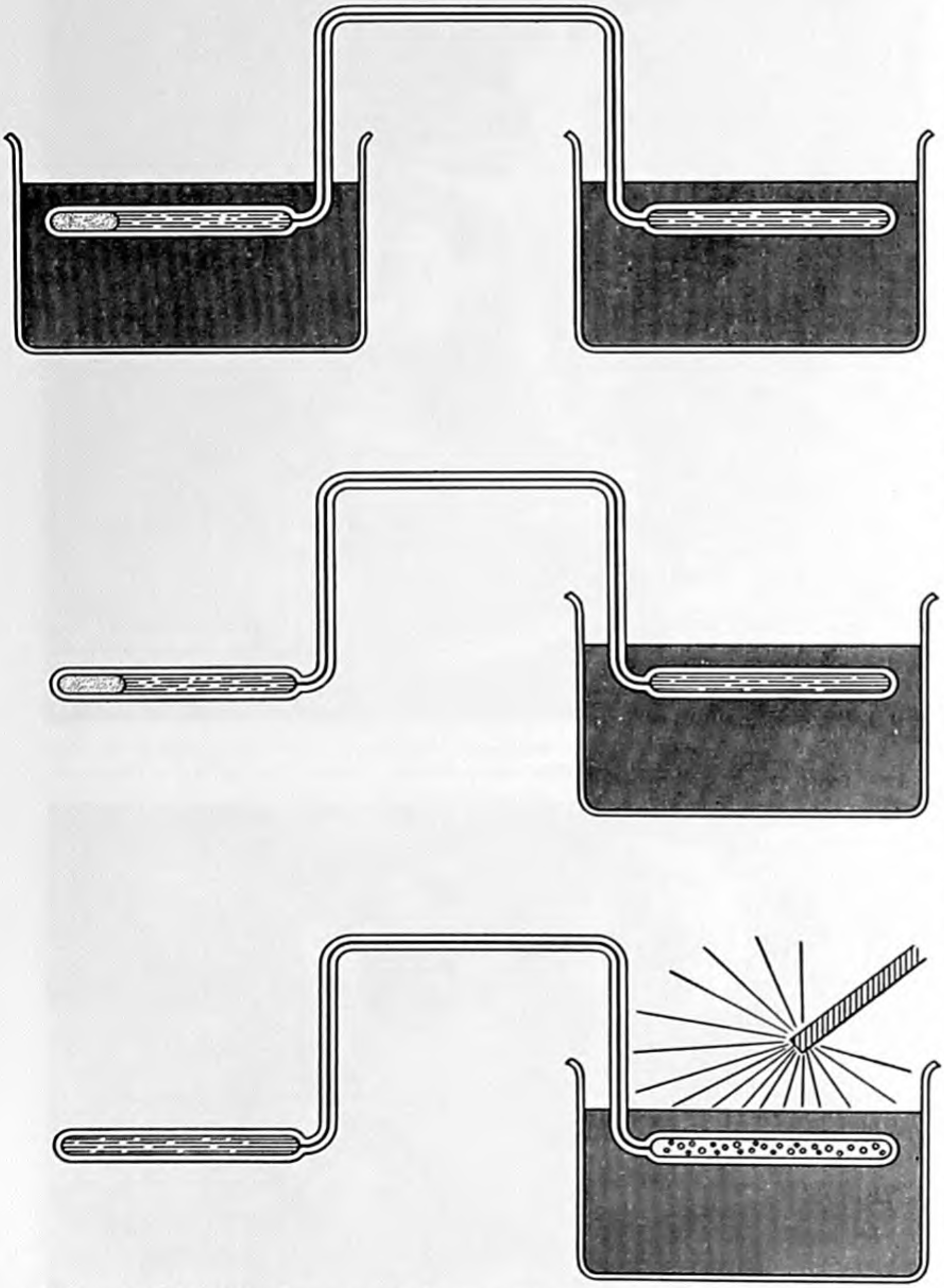
The next step was to find out whether the process could be localized so as to form a track of bubbles. This time I used a somewhat larger bulb (half an inch in inside diameter) filled with ether; it was connected by a capillary tube to a piston-fitted cylinder with a hand crank which could quickly lower the pressure. High-speed movies, at the rate of 3,000 pictures per second, were made of the happenings in the bulb when the pressure was reduced. Sure enough, the pictures disclosed a track of tiny bubbles when a particle darted through the superheated ether. The experiment also demonstrated that the larger the chamber, the shorter the time it can maintain its superheated condition, because of the greater likelihood of intercepting a stray bit of radiation. This bulb remained quiet for only a few seconds after the pressure was lowered.

The bubble type of chamber soon proved to be a very sensitive recorder. Even fast mu mesons, which ionize only lightly, made visible tracks in the superheated liquid. The chamber was placed under four inches of lead and between two Geiger counters connected by an electronic coincidence circuit, so that whenever a particle passed through both counters (and of course the chamber between), the firing of the counters set off a flash, illuminating the apparatus for a camera whose shutter was left open in the darkened room. The purpose of the lead screen and the twin counters was to restrict the photographed events to cosmic rays, and the particles were known to be mu mesons.

By delaying the light flash to go on at a given time after the counters fired, we were able to photograph the bubbles at any stage of their growth. The bubbles grow to a tenth of a millimeter in diameter within a few microseconds after the particle has passed through the liquid. We also found that the number of bubbles along a given length of track depended sharply on the temperature.

The tracks of mu mesons are not very interesting, because these particles rarely break up atomic nuclei. We therefore sought a way to make the instrument photograph other events. When bubbles erupt in the superheated liquid, they emit a distinct "plink" sound. To use this sound as a trigger for the light flash, David C. Rahm, in our laboratory at the University of Michigan, clamped a phonograph pickup to the wall of the bubble chamber and connected it to an amplifier which flashed the light whenever the phonograph needle sensed a vibration. The method did not, however, yield sharp pictures, because sound travels so slowly that the bubbles in the chamber had grown too large by the time the picture was made.

At the University of Chicago R. H. Hildebrand and his colleagues built a chamber like the one just described ex-



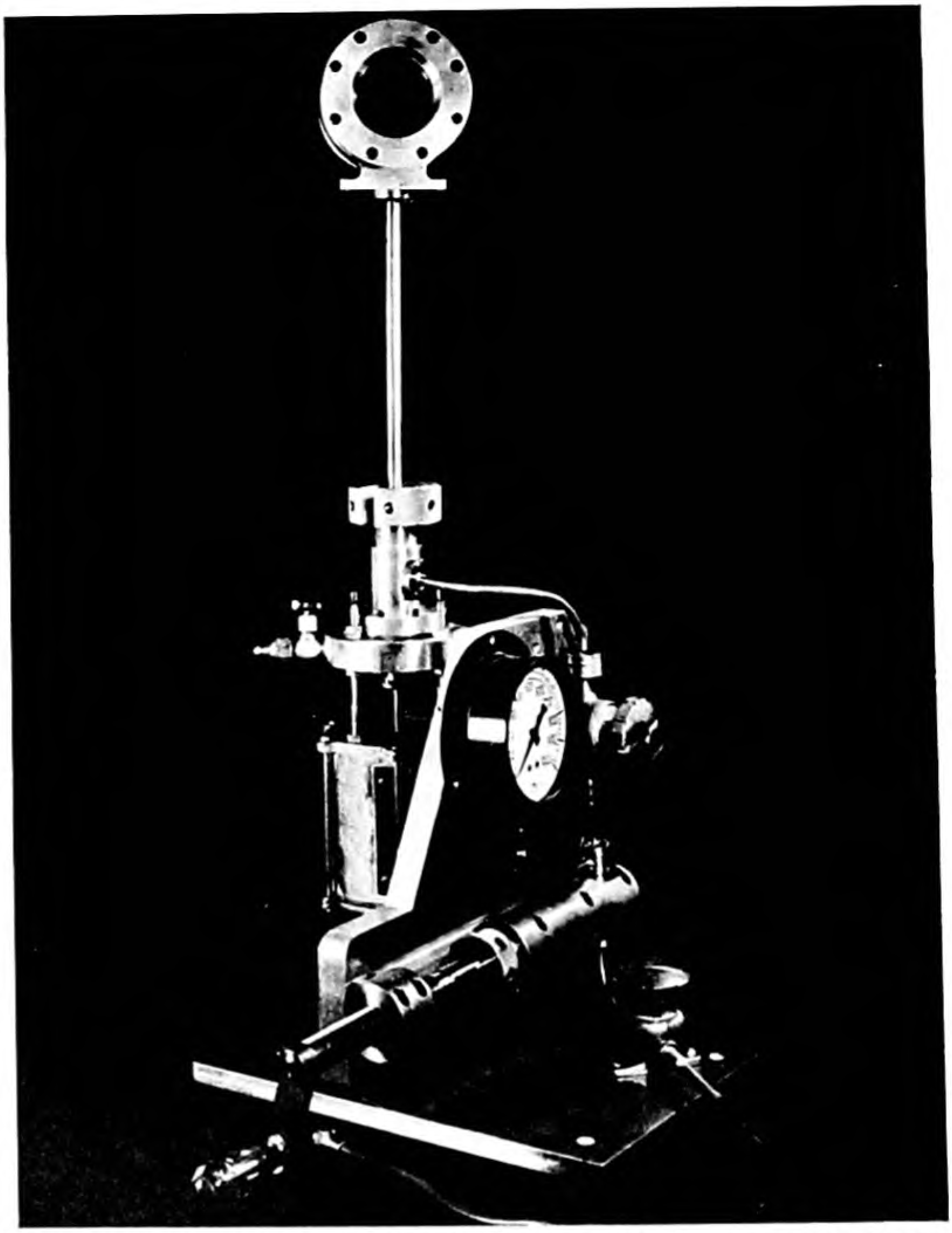
FIRST EXPERIMENT demonstrated that particles can trigger boiling in a superheated liquid. When the left-hand, 320-degree bath was removed, the pressure in the sealed tubes fell from 21 atmospheres to one atmosphere and liquid ether in the 285-degree bath at right became superheated. The ether boiled when it was exposed to radioactive cobalt.

cept that they filled it with liquid pentane instead of ether. They exposed it to a beam of pi mesons from the Chicago synchrocyclotron. Their very beautiful pictures show that bubble tracks may be light or heavy, depending on the type of particle that produces them, and that the density of bubbles along a track is an index of the ionizing power of the responsible particle. Their chamber can be compressed and expanded once a second. It should be able to collect data very rapidly on the interaction of pi mesons with the nuclei of hydrogen and carbon atoms in the pentane.

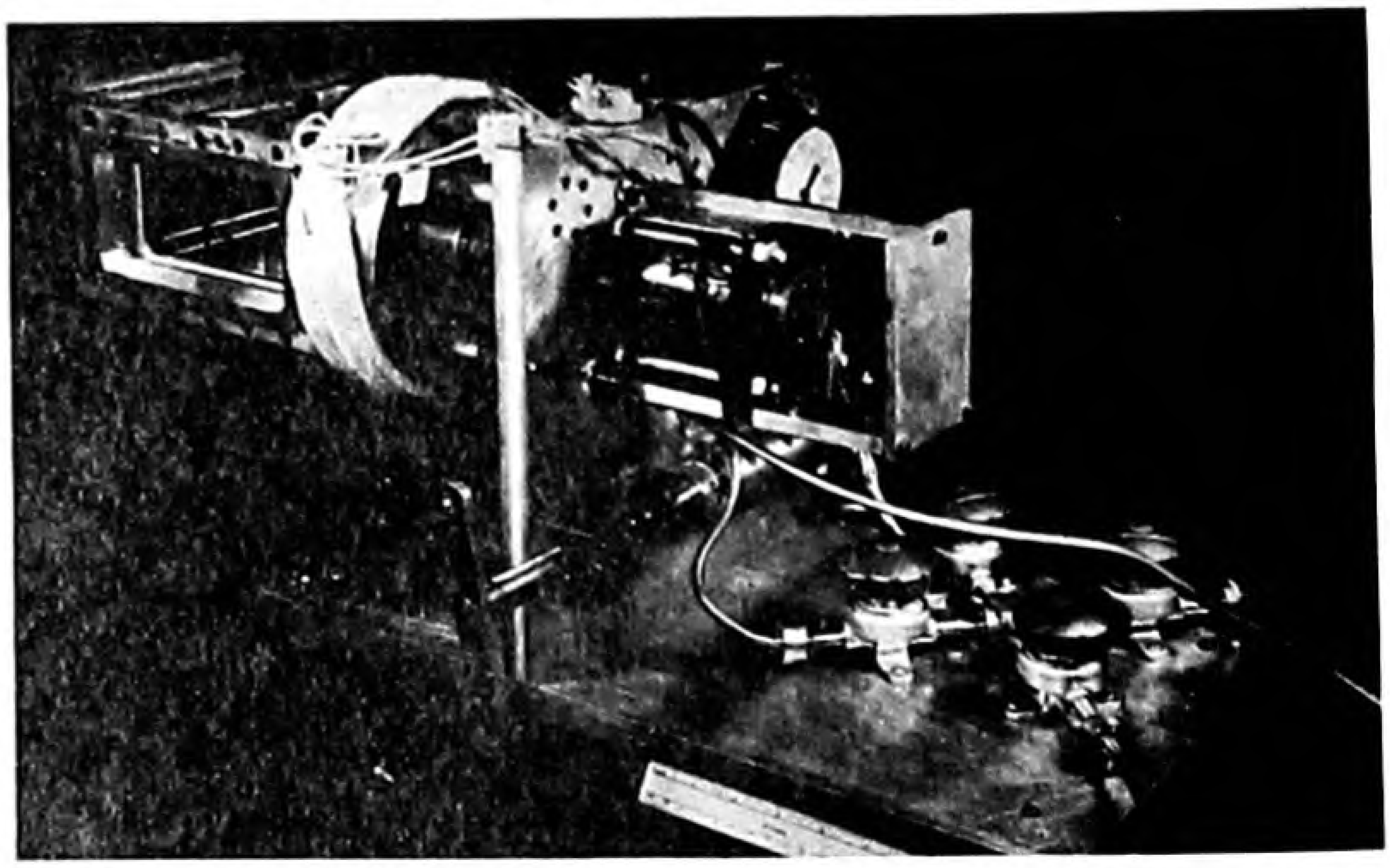
Having demonstrated that the bubble chamber idea worked, we proceeded to the task of building one large enough for practical laboratory use. Flat glass windows were needed to make good pictures. It turned out to be impossible to make an all-glass chamber strong enough to withstand the necessary pressure changes. It therefore had to be made of metal, with small glass windows. However, this introduced a serious problem: rough spots on the metal and at the joints caused the liquid to start boiling. The rise in pressure due to the release of vapor by the boiling at the wall counteracted the reduction of pressure when we expanded the chamber, so that it was impossible to attain the desired degree of superheating. We found that we could get around this difficulty by expanding the chamber very quickly. In other words, the volume of the chamber was increased faster than vapor was evolved at the wall. The desired drop in pressure and superheated condition was maintained for several thousandths of a second.

We first built a two-inch chamber of duralumin and glass, with a diaphragm, actuated by compressed air, which could fully expand the chamber in five thousandths of a second. The liquid remained sensitive for seven thousandths of a second. We then incorporated the same design features in a larger pentane-filled version in which the liquid volume is six inches long, two inches wide and three inches high. This chamber is now in use with the Cosmotron at the Brookhaven National Laboratory. We have made 400 excellent pictures of tracks of protons from this accelerator. Two of the pictures appear on the next page.

These track photographs are as easy to read as the best cloud chamber records and are about 10 times as accurate. The bubbles grow so fast that the tracks can be photographed before swirling motions in the liquid distort them. The



TWO-INCH CHAMBER is made of duralumin and glass. It connects through the thin pipe with a magnetically-operated valve which reduces the pressure in five milliseconds.



LARGEST BUBBLE CHAMBER built so far, measuring six by three by two inches inside, is being tested at Brookhaven National Laboratory. The glass window is at the left.

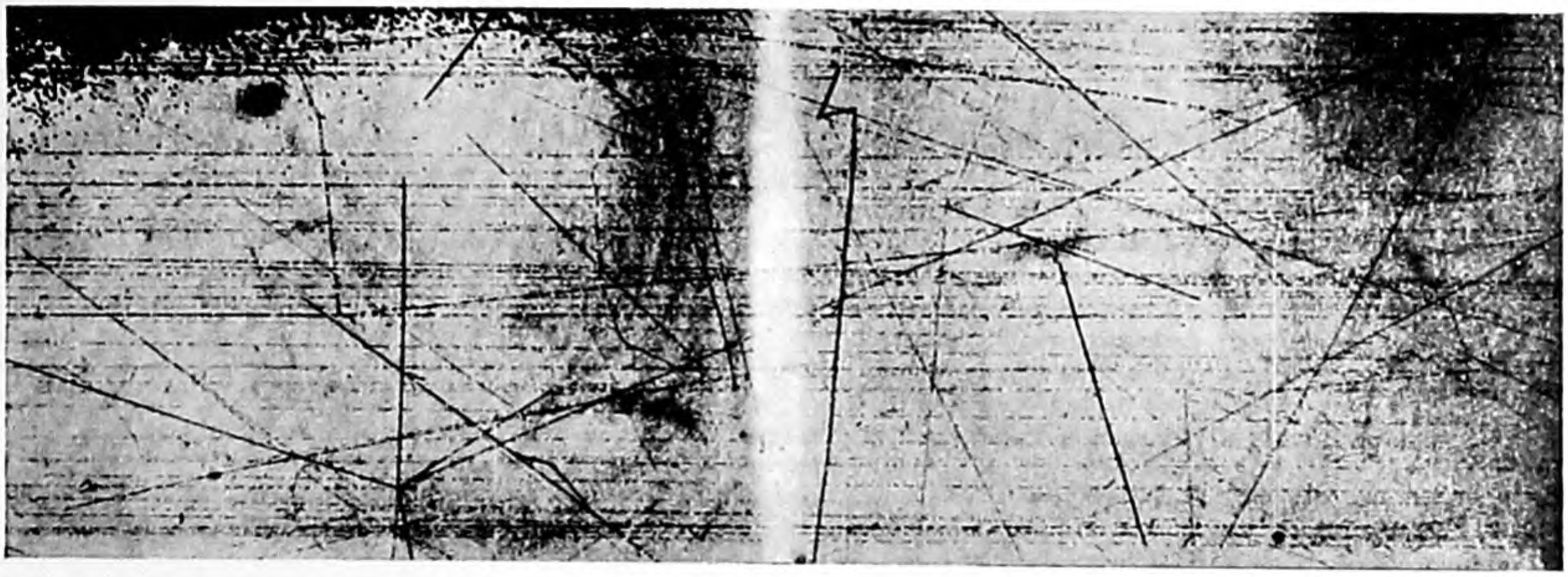
trails in cloud chambers are usually considerably distorted by convection currents. Also, the density of bubbles in a bubble-chamber track may be a better index of particle energy than the droplet density in a cloud chamber.

But the most dramatic difference is in the frequency with which interesting events are recorded. The six-inch bubble chamber catches as many of these as would an ordinary cloud chamber 140 feet long! A typical series of such events is shown at the left of the lower photograph below. A pi meson entered the chamber from the left, was stopped in the liquid and decayed to a mu meson; the latter, after traveling about three millimeters in the liquid, stopped and decayed into an electron. Such complete records of pi-mu-electron decays show up rarely in nuclear emulsions; in cloud chambers they are virtually unknown. (The path of a mu meson in a cloud chamber would be more than three yards long!) Yet the first 22 bubble-chamber pictures from Brookhaven, taken in 11 minutes, contained eight of these events. Among other interesting features in the pictures reproduced here is the destruction of a carbon nucleus with the creation of a pair of pi mesons. One especially inviting puzzle that still awaits analysis is the strange zigzag track down the middle of the upper photograph on this page.

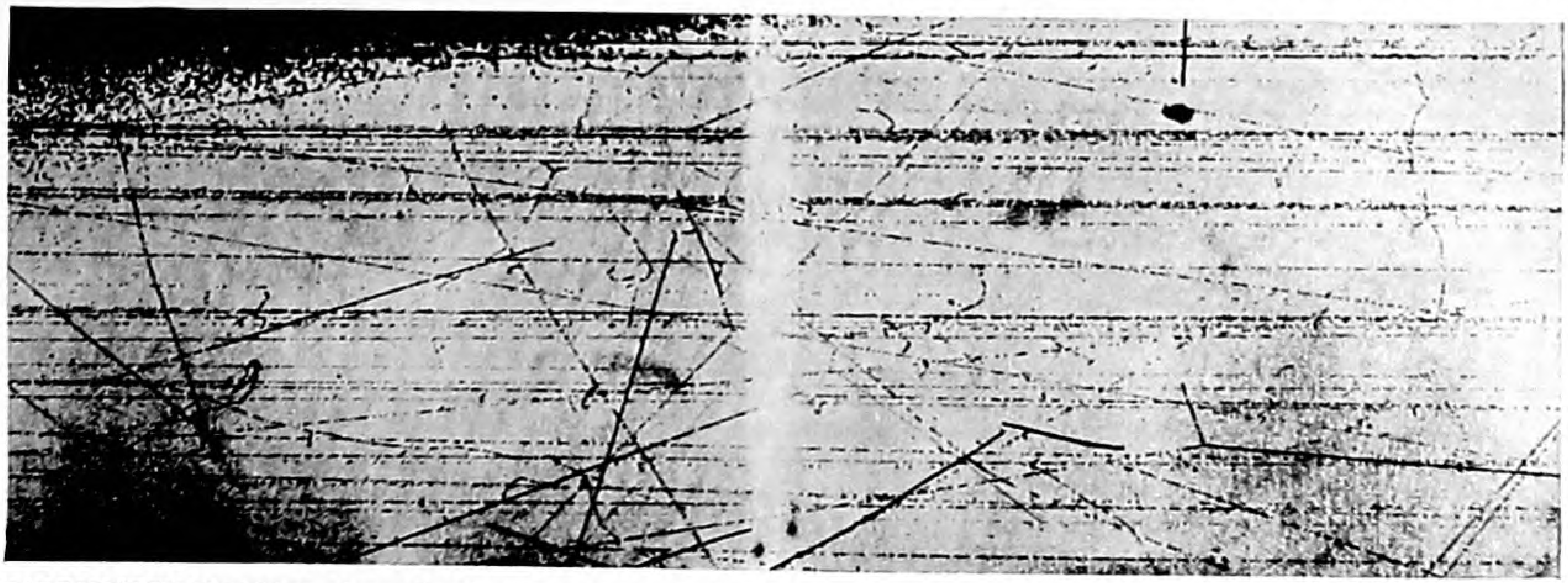
Another advantage of the bubble chamber is that it may be filled with a light-atom liquid, which does not deflect particles much and therefore will permit magnetic-field experiments, or with an extremely dense liquid, which will produce a great amount of scattering, as an emulsion does.

Often nuclear physicists wish to bombard a single "elementary" particle—*e.g.*, protons—to see what fragments may be produced. They direct the bombarding beam at liquid hydrogen as the target and observe the results by means of detectors placed around it. The bubble chamber makes it possible to have the target material within the chamber itself. Superheated liquid hydrogen will boil and show particle tracks. Luis W. Alvarez and his group at the University of California have made pictures of protons in the liquid hydrogen recoiling from collisions with incoming neutrons [*see photograph on page 109*].

It appears that the bubble chamber will become a standard detection instrument for work with the high-energy particle accelerators in laboratories, and that it will speed up the rate at which we can gain information about strange particles and nuclear forces.



BUBBLE-CHAMBER PICTURE shows the two-billion-electron-volt proton beam of the Cosmotron at the Brookhaven Laboratory. The protons enter the chamber from the left. At lower left a carbon nucleus has been disrupted. Zigzag track in the center is a mystery.



RARELY SEEN EVENT, the decay of a pi meson into a mu meson and then into an electron, was caught in this photograph. Pi meson enters just above center of left edge, moving down to right. Mu meson leaves a short track and electron goes off upward to left.

The Author

DONALD A. GLASER is professor of physics at the University of Michigan. Born in Cleveland in 1926, he went to the Case Institute of Technology with the idea of becoming a mechanical engineer. After six weeks he decided against engineering and turned to physics. He went to do his graduate work at the California Institute of Technology under Carl D. Anderson, and for his Ph.D. thesis investigated high-energy cosmic rays. Glaser claims that the encouragement to try his bubble-chamber idea, which he had already been mulling over, came one night during a beer session with some physicist friends: "After several pitchers of beer we began to wax philosophical about physics. One of the boys, looking dreamily into the pitcher of beer before him, saw the usual streamers of bubbles and remarked, 'Nuclear physics should be easy. You can see tracks in nearly everything.' Just for fun I actually exposed some beer to gamma rays the next day in the laboratory. Nothing happened." But as he explains in his article, he went on to more serious tests which succeeded. Glaser is a devotee of chamber music; while in high school he was a violist in the Cleveland Philharmonic Orchestra, and he still joins friends in quartets.

Bibliography

CHARACTERISTICS OF BUBBLE CHAMBERS. D. A. Glaser and D. C. Rahm in *The Physical Review*, Vol. 97, No. 2, pages 474-479; January 15, 1955.

PROGRESS REPORT ON THE DEVELOPMENT OF BUBBLE CHAMBERS. D. A. Glaser in *Nuovo Cimento*, Vol. 11, Series 9, Supplement No. 2, pages 361-368; 1954.

SOME EFFECTS OF IONIZING RADIATION ON THE FORMATION OF BUBBLES IN LIQUIDS. Donald A. Glaser in *The Physical Review*, Vol. 87, No. 4, page 665; August 15, 1952.

SCIENTIFIC AMERICAN December 1958, Vol. 199, No. 6, pp. 42-50 OFFPRINT 215

THE MASER

by James P. Gordon

It is a quantum-mechanical device which amplifies very short radio waves with extraordinary fidelity. The basic element of the early masers was gaseous ammonia; the latest models make use of crystals.

The 50-foot radio telescope at the Naval Research Laboratory in Washington, D.C., recently acquired a strange accessory. Mounted just behind the antenna, at the center of the telescope's parabolic reflector, is an oblong box containing a synthetic ruby and some standard microwave equipment. A bath of liquid helium chills the ruby to the temperatures of the cold reaches of space which the telescope surveys. With the help of this refrigerated gem astronomers hope to extend their range of observation far beyond its present limits, perhaps far enough to clear up once and for all the mysteries of the size and geometry of the universe.

The ruby is part of a new microwave device called the "maser." The letters of this odd word stand for Microwave Amplification by Stimulated Emission of Radiation. The maser represents the ultimate in high-fidelity amplifiers. The best previous amplifiers, using vacuum tubes, put out a mixed signal which combined an amplified version of the input with a wide assortment of oscillations originating in the tubes themselves. If the input signal becomes weaker, the percentage of noise in the output increases, and the resemblance between input and output diminishes. Eventually a point is reached where the input, though still amplified, can no longer be recognized in the output. The great virtue of the maser is that it generates practically no noise. It can detect much weaker signals than other amplifiers can, and hence pick up radio waves from far more distant points in space. As we shall see, it is also finding a number of other important applications in science and technology.

What makes the maser so quiet? It is perhaps helpful to ask first: What makes vacuum tubes so noisy? Vacuum tubes utilize a stream of agitated electrons which are boiled out of a cathode and sent crashing into a collecting plate by an outside voltage. The signal to be amplified imposes its variations on the electron stream. But the particles have their own random variations, which are inevitably part of the output of the tube. It is a tribute to the ingenuity of electrical engineers that, in improvements such as the traveling-wave tube, they have been able to go so far toward muffling the effects of unruly electrons. The least noisy of these tubes, however, leaves a lot to be desired.

The maser dispenses with streams of electrons altogether. Instead it makes use of certain intrinsic oscillations in many types of material particles. These oscillations are basic phenomena of nature. The idea of harnessing them for useful work occurred independently a few years ago to several workers in the field of microwaves, including C. H. Townes of Columbia University, N. G. Basov and A. M. Prokhorov in the U.S.S.R. and J. Weber of the University of Maryland.

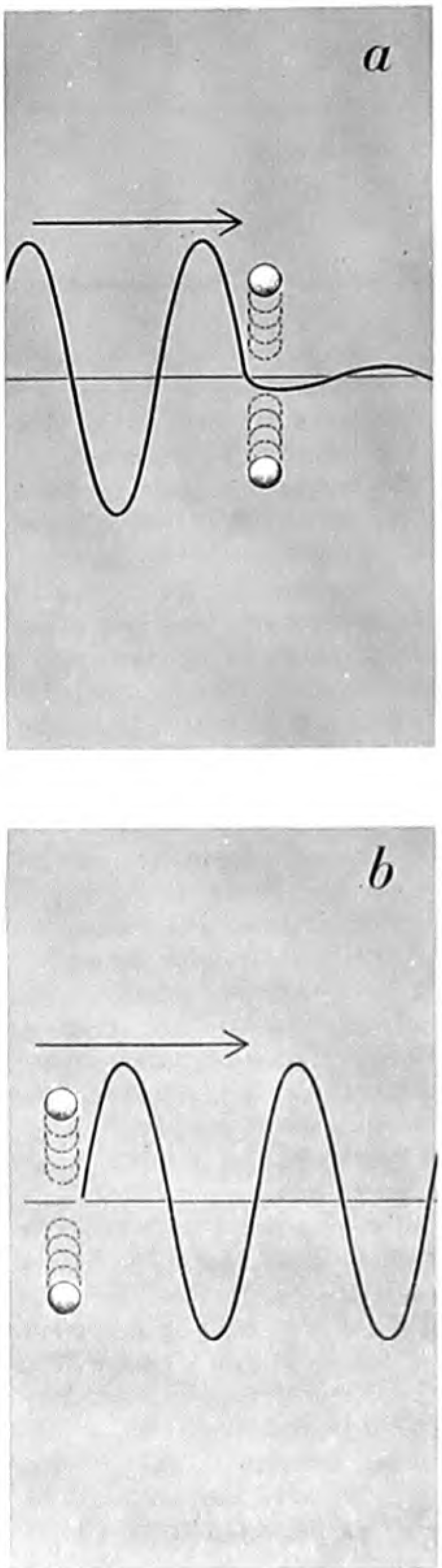
To appreciate what led to this notion we should briefly consider the interaction between high-frequency radiation and matter. Every student of elementary physics has witnessed the experiment in which light from a sodium lamp is shined into a container of cool sodium vapor and is completely absorbed. At the same time light of a different frequency—that is, color—from some other source passes through the container undimmed. The classical explanation is that every atom and molecule has certain natural vibrations which occur at sharply defined frequencies. When the oscillations of light or of other electromagnetic waves coincide with one of these frequencies, the radiation gives up energy to the atom or molecule, causing it to vibrate like a pendulum which has been set swinging by a series of properly timed pushes. Conversely, if atoms or molecules can be made to vibrate by some other means, say by thermal agitation, they will emit electromagnetic waves of the same characteristic frequency. In the experiment just mentioned, waves from hot, vibrating sodium molecules are absorbed by the cold molecules. Waves whose frequency does not correspond to the frequency of the sodium vibrations pass through unaffected.

If the reader is wondering how such a mechanism can be made to amplify the energy in a wave, he may as well stop. If the "classical"—that is, pre-quantum mechanical—explanation were completely correct, there would be no maser. (As a matter of fact, there would be no atoms. On the classical theory electrons revolving around atomic nuclei would continuously radiate away their energy and spiral into the nucleus. All of ordinary matter would thus collapse.)

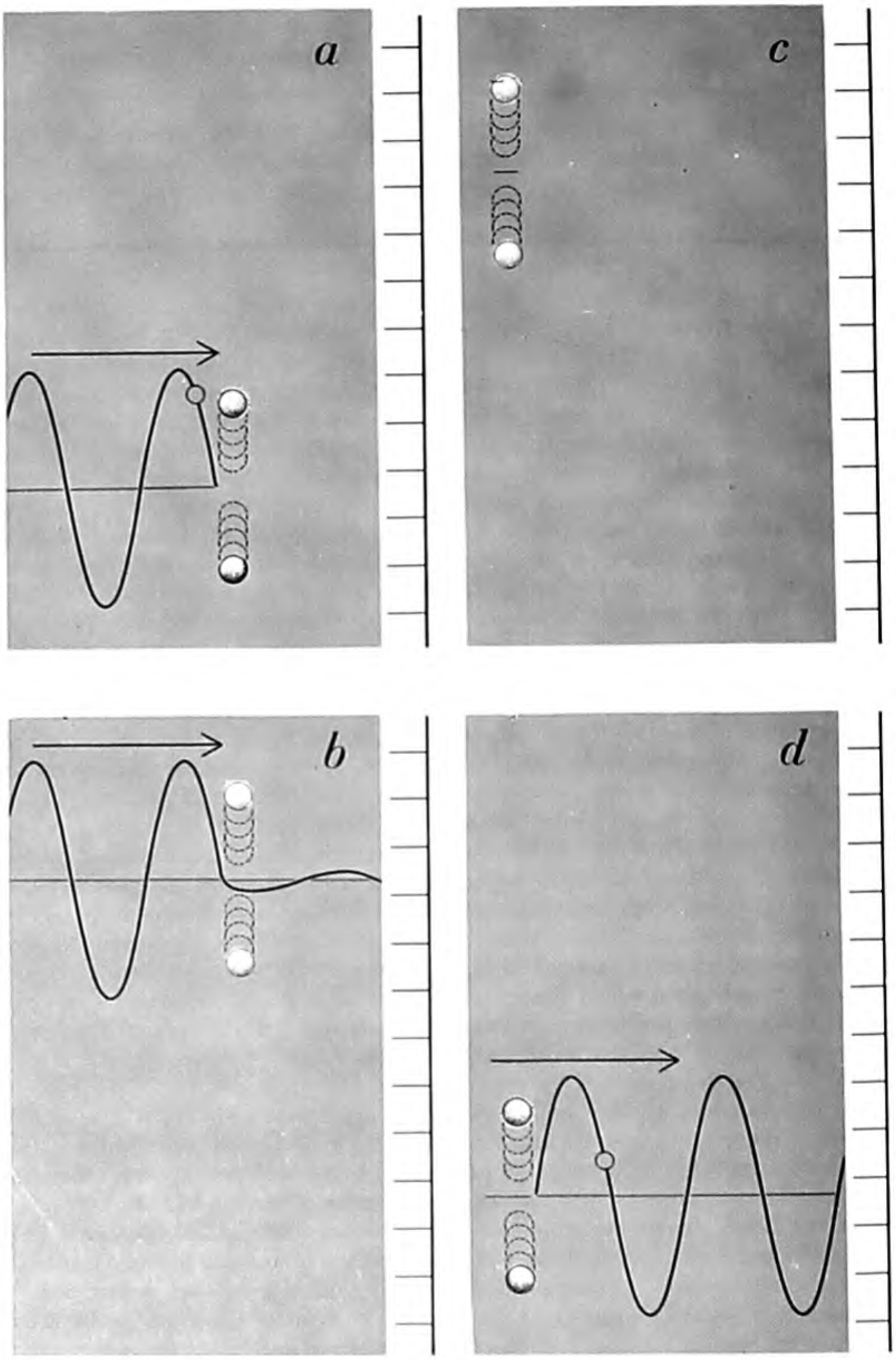
To discover the secret of the maser we must turn to the quantum picture of matter and radiation. In this view atoms and molecules exist most of the time in one of a number of stable, nonradiating states. Each state corresponds to a fixed quantity of energy. Radiation, on the other hand, consists of the particles called photons, carried by a sort of guiding wave. The frequency of the wave is a measure of the energy of the photons, according to Max Planck's famous equation $E = hf$. A particle of radiation is produced when an atom falls from a higher to a lower energy state, and the energy of the photon is exactly equal to the difference in energy between the states. When an atom jumps the other way, from a lower to a higher energy state, it absorbs a photon of the same frequency. Thus

when radiation passes through an assembly of atoms, one of three things can happen. If the energy of the photons does not equal the difference between a pair of energy levels in the atoms, there is no interaction. If the energies match, and a photon collides with an atom in the lower of the two states, the radiation will be absorbed and the atom will be "excited" to the higher state. If the photon collides with an atom in the higher state, it will cause the opposite jump, down to the lower state, and a new photon will be emitted. Thus there will now be two photons where before the collision there was only one.

In any assemblage of atoms there is always some traffic between low- and high-energy states. The atoms keep hopping up and down in their energy states, boosted by energy received in chance collisions and falling because of their natural tendency to seek the lowest energy level. Under ordinary conditions the lower states are always more densely populated than the higher ones. Thus when radiation of the appropriate fre-



CLASSICAL VIEW of the interaction of electromagnetic radiation and matter is depicted. At top an electromagnetic wave (*wavy line*) of the appropriate frequency sets a two-atom molecule (*balls*) to vibrating. The process absorbs energy from the wave. At bottom a molecule spontaneously emits energy (*wavy line*) of characteristic frequency.



QUANTUM-MECHANICAL VIEW of the interaction is similarly depicted. Here the electromagnetic radiation is regarded not as a wave but as a photon (*white dot*) guided by a wave. The frequency of the guiding wave is related to the energy of the photon. The molecule does not vibrate; the broken circles merely indicate that the atoms are regarded as simultaneously occupying a number of positions. At top left the molecule is at a lower energy level. At bottom left the molecule has been "excited" by a photon of the appropriate energy, and raised to a higher energy level. At top right the molecule is at the higher energy level. At bottom right it has fallen to the lower energy level and emitted a photon of characteristic energy. Scale at right of these four illustrations suggests energy levels.

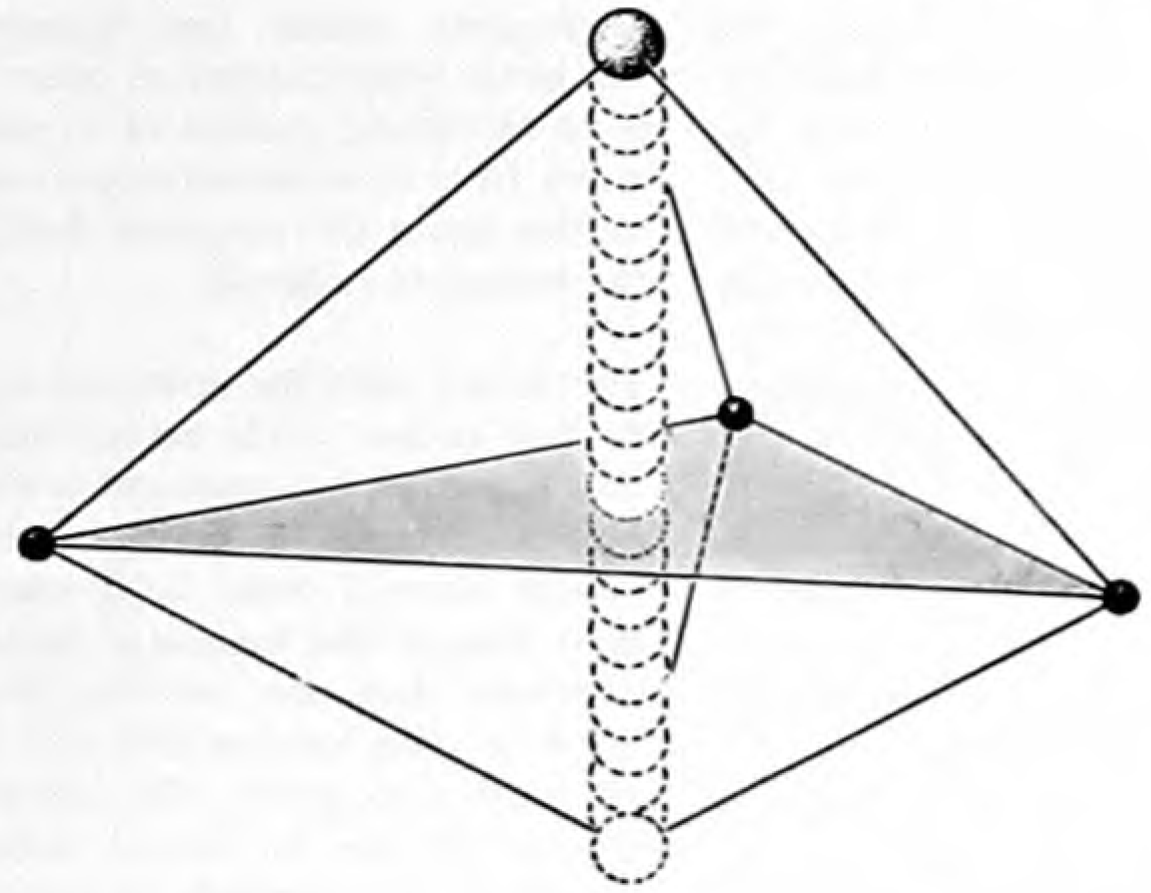
quency passes through the assemblage of atoms, more photons will be absorbed than new ones created, and the outgoing beam will be weaker than the incoming beam.

But suppose it were somehow possible to change the distribution of energy levels so that there were more atoms in the higher of two states than in the lower. Then a beam of photons of the appropriate frequency would produce more downward jumps than upward ones; the net effect would be that more photons would come out than went in. In other words, the output wave would have more energy than the input wave. This is the secret of how the maser amplifies.

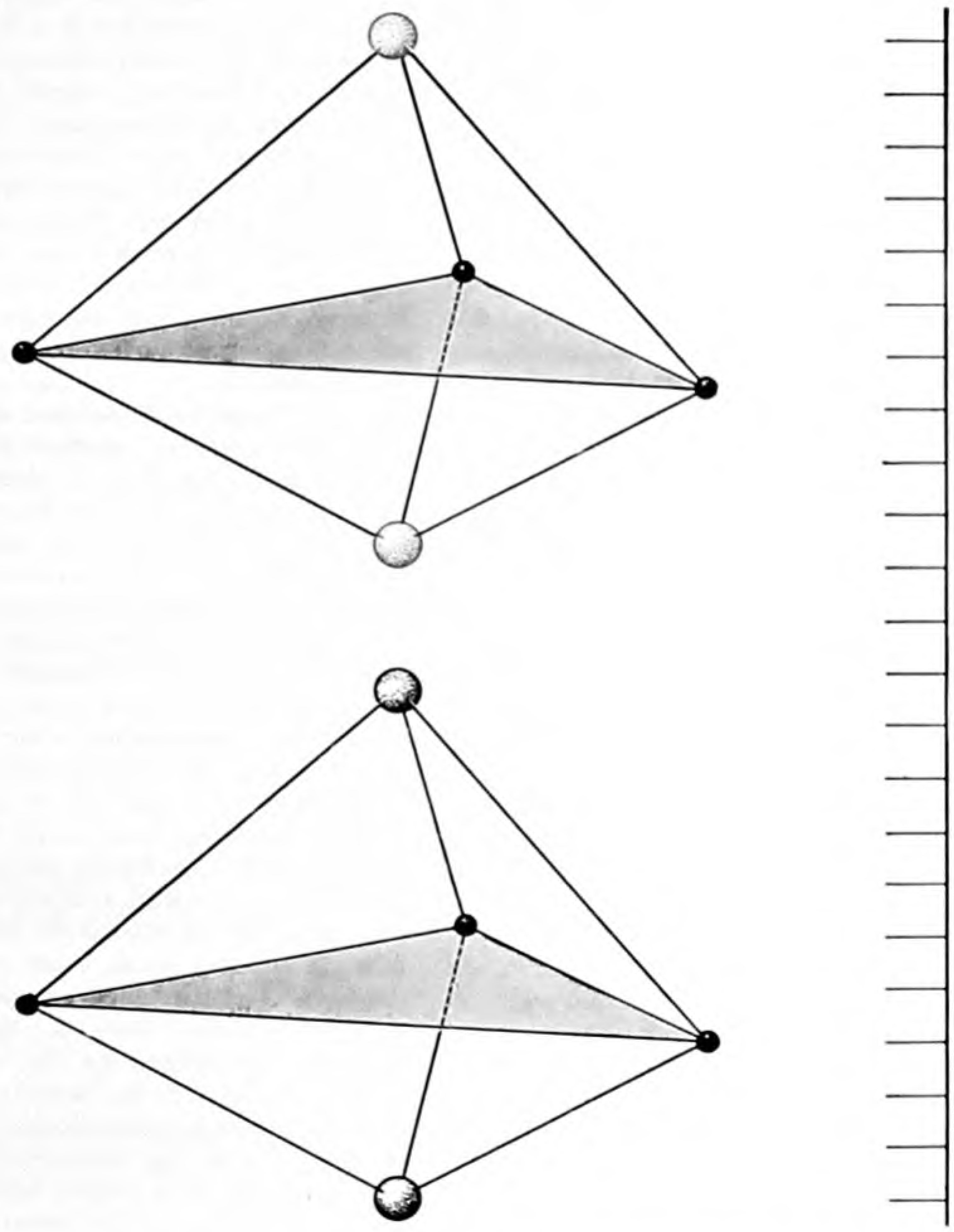
The first of the masers to be developed is based on the molecule of ammonia. For reasons we shall mention in a moment, the ammonia maser is more useful as an oscillator and a timekeeper than as an amplifier. It is in any case a remarkable device: a simple metal chamber, into which only a little ammonia gas is admitted, which yields a weak microwave signal of almost unbelievable purity. Its output wave falls short of a mathematically perfect sine curve by less than one part in 100 billion!

The molecule which produces this perfect monotone has the shape of a pyramid. At the apex of the pyramid is a nitrogen atom; at the base, three hydrogen atoms [*see illustrations at right*]. The nitrogen atom is able to move through the plane of the hydrogen atoms. thus turning the pyramid inside out. On the classical theory we picture the nitrogen atom flipping back and forth at a characteristic frequency of about 24,000 million vibrations per second, or 24,000 megacycles per second. At any given instant the nitrogen atom is on one side of the hydrogens or on the other. From the quantum point of view the nitrogen has at a given time a certain probability of being on either side—in a sense it is partly on both sides. Moreover, the molecule as a whole has two distinct energy states. The difference in energy between the states equals the energy of a photon with a frequency of 24,000 megacycles per second.

Now it happens that ammonia molecules in the higher state are repelled by strong electrostatic fields, whereas those in the lower state are attracted. Thus we have a method for segregating the high-energy molecules and getting maser action. The separator is a cylinder of charged rods [*see illustration at top of page 120*]. In the vicinity of the rods the field is strong; along the cylinder's axis the field is weak. When a beam of



CLASSICAL VIEW OF THE AMMONIA MOLECULE is that its single nitrogen atom (*large ball*) vibrates back and forth across the plane of its three hydrogen atoms (*small balls*).



QUANTUM-MECHANICAL VIEW OF THE MOLECULE is that, in a sense, the nitrogen atom is simultaneously on both sides of the plane of the hydrogen atoms. The molecule may occupy either a higher energy level (*top illustration*) or a lower energy level (*bottom*).

ammonia molecules is sent through the separator, those in the upper state are attracted to the axis, while those in the lower state are pulled toward the electrodes and dispersed. Out of the far end comes a stream of molecules, virtually all of which are in the upper energy-state. If these molecules are irradiated with 24,000-megacycle microwaves, only downward transitions will be induced. Energy will be given up by the molecules to the microwave field, and the incoming wave will be amplified.

In the actual instrument ammonia gas at low pressure escapes from a nozzle into an evacuated chamber containing the separator and a resonant cavity. After passing through the separator, molecules in the upper state enter the cavity, into which the microwave signal is fed through a waveguide.

The resonant cavity is simply a metal box with highly reflecting walls. Each incoming photon can bounce back and forth across the chamber thousands of times before it escapes again, greatly increasing its chance of interacting with a molecule in the beam.

Whenever there is a collision, a new photon is born. It too is trapped in the chamber for a time and may collide with another molecule, producing a second new photon, and so on. If there are enough molecules in the cavity, this chain reaction becomes self-sustaining; the amplifier turns into an oscillator, generating its own wave without any input signal.

The ammonia maser is an extraordinarily stable oscillator. Its virtually unvarying sine waves can be used as a "pendulum" to regulate an almost perfect clock [see "Atomic Clocks," by Harold Lyons; SCIENTIFIC AMERICAN Offprint 225]. Although such timepieces have not yet been fully tested, it has been demonstrated that two ammonia masers will maintain their frequencies with respect to each other for at least a year with an accuracy of one part in 10 billion. A maser-regulated clock should gain or lose no more than one second in a few hundred years.

As an amplifier the ammonia maser has a remarkably narrow band-width: it will not amplify waves which depart from its central frequency by more than 3,000 to 5,000 cycles. The ammonia maser is not readily tunable; the central frequency cannot easily be changed. This means that it is not really a practical amplifier. If it were used in a communications channel, it could transmit only one voice at a time; it could not come close to receiving a television station. The ammonia maser was, however, the instrument which first demonstrated the great potentialities of maser amplifiers. Moreover, studies of its resonance curve have contributed important information about the magnetic fields within the ammonia molecule.

It was not until the invention of masers that utilize solids rather than gases that practical low-noise microwave amplifiers became a reality. Solid-state masers have a noise level even lower than that of the ammonia maser. Furthermore they are tunable, they have much broader band-widths and they put out much more power. The fact that their frequencies can be varied makes them unsuitable as standards of frequency or of time, but it adds considerably to their general usefulness as amplifiers.

The action of the solid-state maser also depends on quantum jumps, but they are jumps of electrons within individual atoms rather than energy transitions of whole molecules. It is by now a familiar fact that every electron is in effect a small spinning magnet. In most atoms, which are nonmagnetic, the electrons are paired off with their poles opposed to each other so that their magnetism is canceled out. There are a few substances, however, in whose atoms the cancellation is incomplete; some electrons are unpaired and the material as a whole is magnetic, or, in technical terms, paramagnetic.

It is the behavior of unpaired electrons placed in an external magnetic field that makes the solid-state maser possible. As usual, there are two ways to describe this behavior: the classical way, which has the advantage of being easy to visualize but the drawback of being incomplete; and the quantum way, which is implausible but correct. Classically we imagine that the spin axis of the electron wobbles, or precesses like a top around the direction of the field [*see diagram at left in illustration at top of opposite page*]. In quantum terms we say that the spinning electron can have just two positions: one in which its axis points in the same direction as that of the field; the other in which it points in the opposite direction. The two positions constitute different quantum states, the higher of which is represented by the electron whose axis points in the direction of the field. As in the case of molecules, the difference between the levels corresponds to the energy of a photon whose frequency equals that of the classical vibration. Also as in the case of molecules, there are normally more electrons at lower levels than at higher.

To make a maser we simply need to find a way of reversing the normal distribution and putting the majority of electrons in the upper state. Then if they are irradiated with photons of the correct frequency, they will jump down, amplifying the incoming beam.

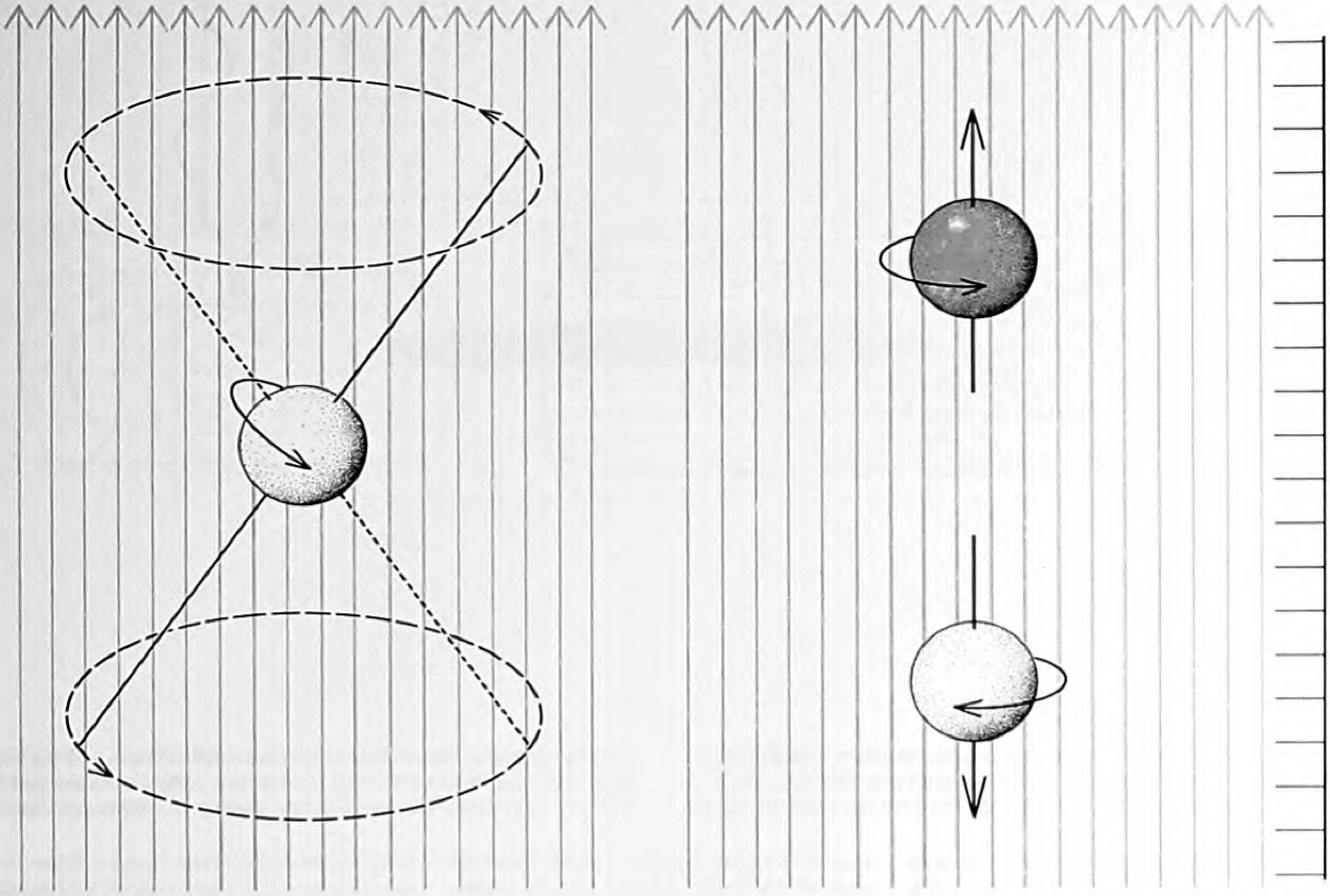
The first type of solid-state maser that was developed is known as the two-level paramagnetic maser. In some versions the paramagnetic material is a silicon crystal containing some impurity atoms, such as those of phosphorus, which have one more electron than they need to satisfy their role in the crystal lattice. In other versions it is a quartz crystal which has been subjected to neutron bombardment to release unpaired electrons. The crystal is placed between the poles of a strong magnet and cooled to a temperature a few degrees above absolute zero in a bath of liquid helium, so that most of its unpaired electrons fall into the lower of their two possible energy states. Then it is subjected to a fairly high-powered microwave pulse, which briefly raises the majority of the electrons to the higher state. While this "inverted population" of electrons lasts, it can act as an amplifier for a weak microwave signal. In silicon the amplifying period lasts about a minute after each "pumping" pulse; in quartz, only a few thousandths of a second.

The difference between the energy of the upper and lower levels depends upon the strength of the magnetic field. Hence by adjusting the strength of the magnet, the maser can be tuned over a wide range of frequencies. With very strong fields it may be possible to reach the never-never land of waves a fraction of a millimeter long.

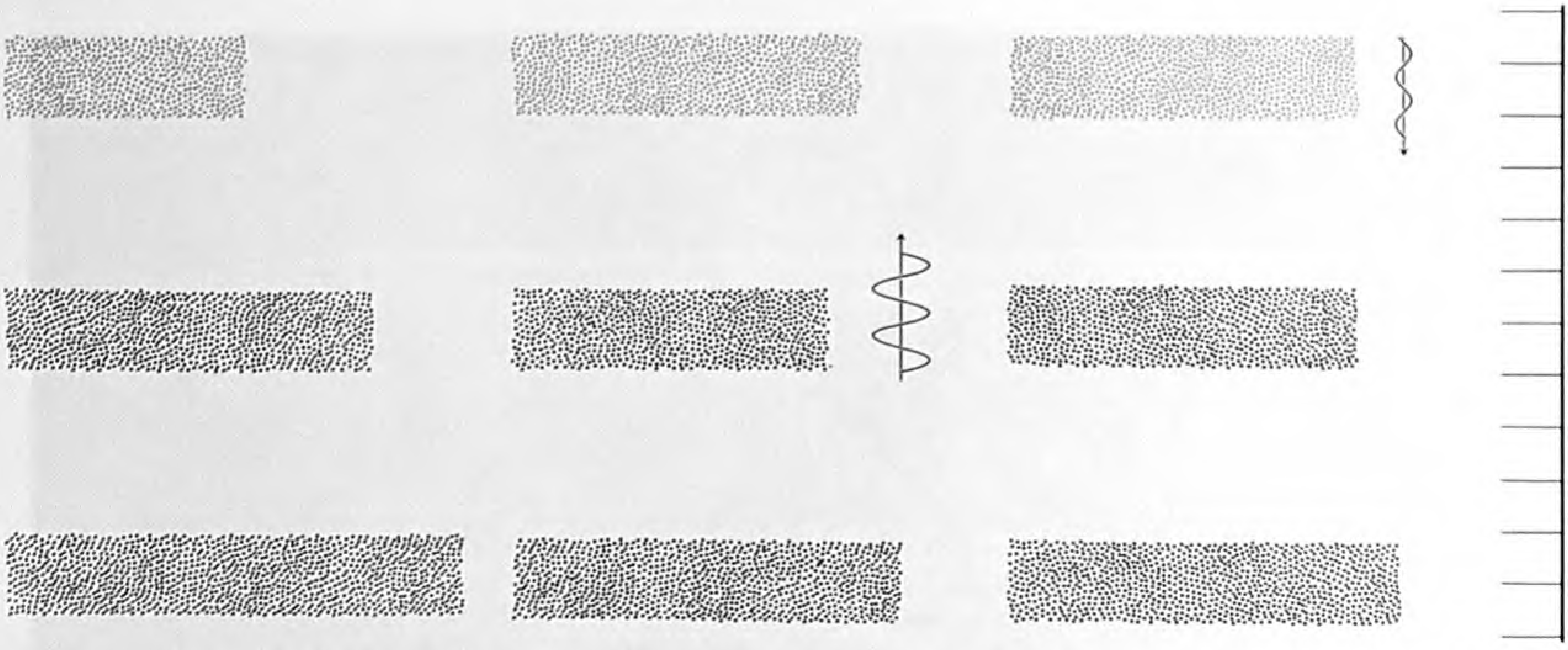
In a solid crystal the outside field is not the only one to act on unpaired electrons. The electrons are also influenced by the magnetism of neighboring atoms. The internal magnetic effect varies from point to point in the crystal, so that not all the electrons are subjected to exactly the same field. Thus they respond to slightly different frequencies, and this is the reason for the wider band-width of the solid-state devices.

The chief disadvantage of the two-level maser is that it can be operated only in bursts; its amplifying action stops each time its electrons drop down again to the lower level. This problem has been overcome with the development of the newest member of the maser family: the three-level paramagnetic maser.

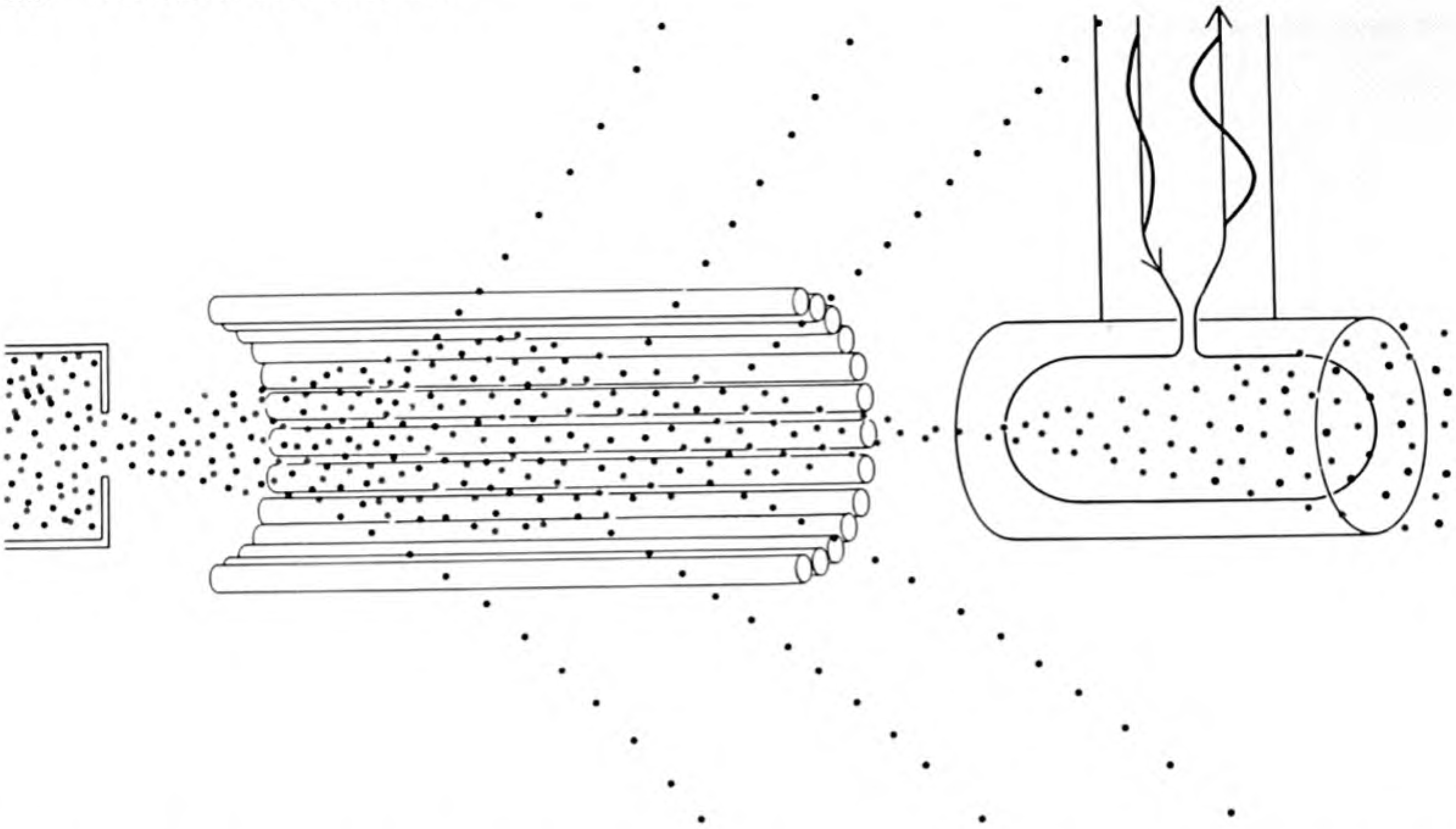
Conceived by Nicolaas Bloembergen of Harvard University, the three-level paramagnetic maser has a basic element



ELECTRON IN A MAGNETIC FIELD *(colored arrows)* is depicted from the classical standpoint *(left)* and from the quantum-mechanical *(right)*. In the classical view the axis of the electron's spin precesses, or wobbles, around the direction of the magnetic field at a frequency related to the strength of the field. In the quantum-mechanical view the electron has a higher energy state *(top right)* in which its "south" magnetic pole is pointed in the direction of the field, and a lower energy state *(bottom right)* in which the pole is pointed in the opposite direction. The difference in the energy levels is related to the strength of the field.



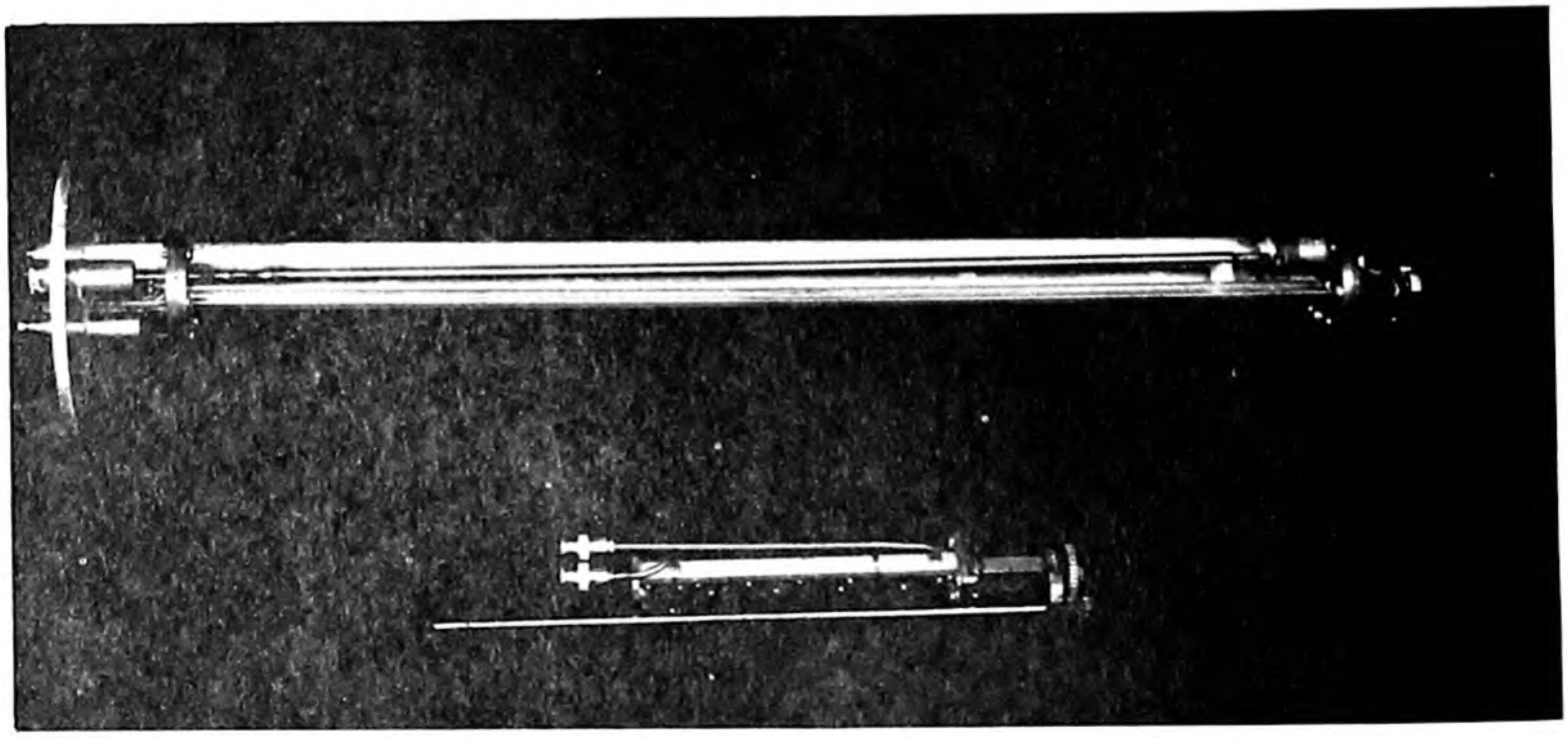
THREE-LEVEL SOLID-STATE MASER is considered. At left are electrons in three energy states; the largest number of electrons is in the lowest state, the smallest number is in the highest state. In the middle electrons are "pumped" from the lowest state to the highest by microwave energy of the appropriate frequency. At right electrons drop from the highest state to the middle state, emitting microwave energy of a lower frequency. Thus energy put into the maser at the latter frequency can be amplified.



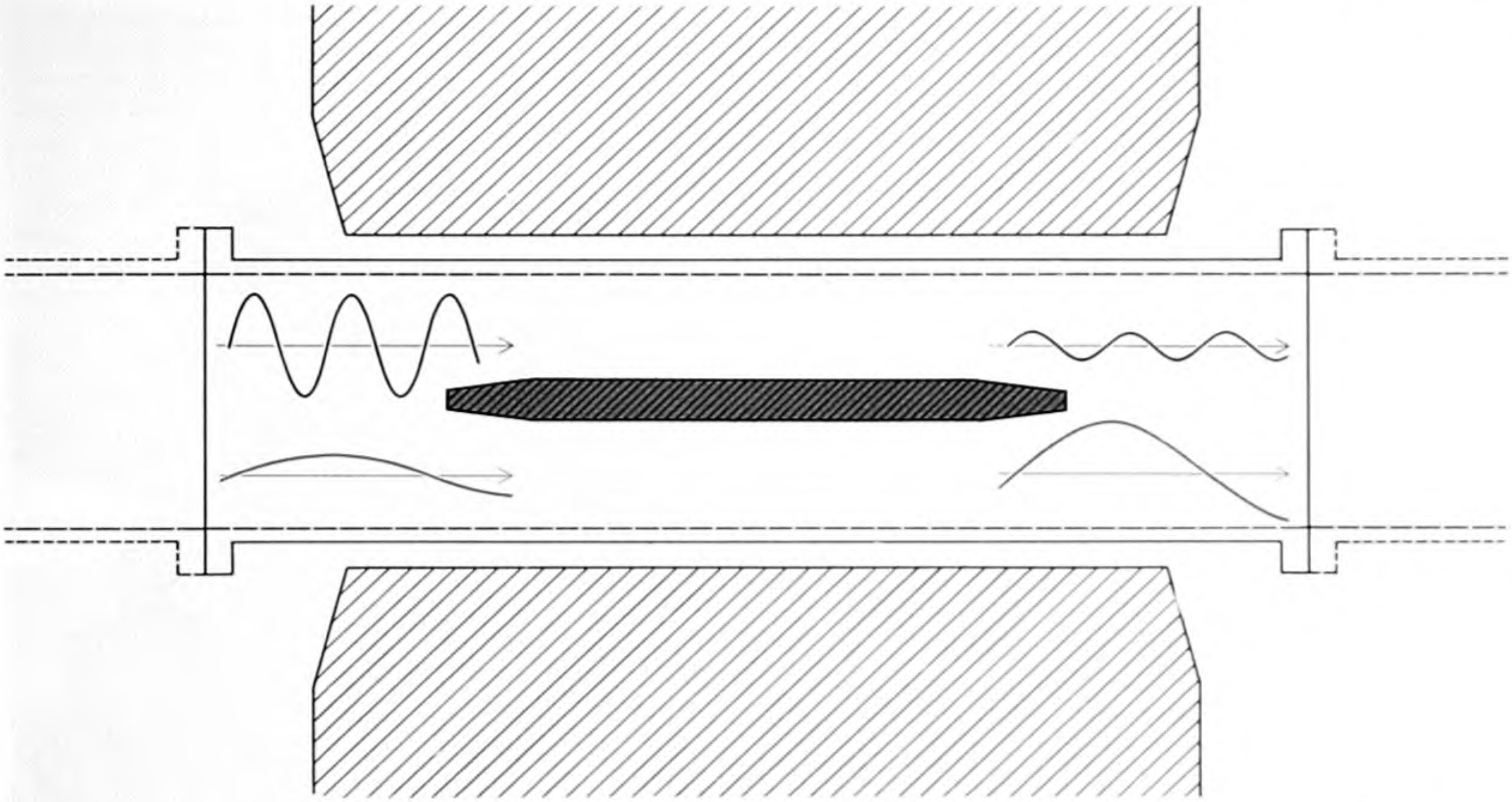
AMMONIA MASER sends ammonia molecules in two energy states through a cylinder of electrically charged rods. The molecules in the lower state (*black dots*) are pulled toward the rods; the molecules in the higher state (*colored*) are attracted to the axis of the cylinder. The molecules in the higher state then enter a resonant cavity (*right*), where they may be used to amplify a microwave signal.

consisting of atoms with more than one unpaired electron apiece. Atoms of this kind are found in the naturally paramagnetic elements such as iron and chromium, in which one of the interior shells of electrons is not filled. Quantum mechanics tells us that in many such atoms there is one more energy level than the number of unpaired electrons. For example, chromium atoms, which make up part of the ruby crystal, possess three unpaired electrons and thus have four energy levels.

Any three of the available levels can be used. When the crystal is cooled to very low temperatures, the atoms distribute themselves among the energy states in the usual way, each higher level containing fewer atoms than the one below it [*see illustration at bottom of preceding page*]. Now we irradiate the solid with microwaves of the proper frequency, causing a jump from the lowest of our three levels to the highest. By this pumping action the top level is kept fuller than the middle one. Therefore a weak signal whose frequency corre-



COMPONENTS OF A THREE-LEVEL MASER appear in the photographs on these two pages. The object at top in the photograph at left is essentially a waveguide through which microwaves are conducted to the maser cell. The object at bottom in the same

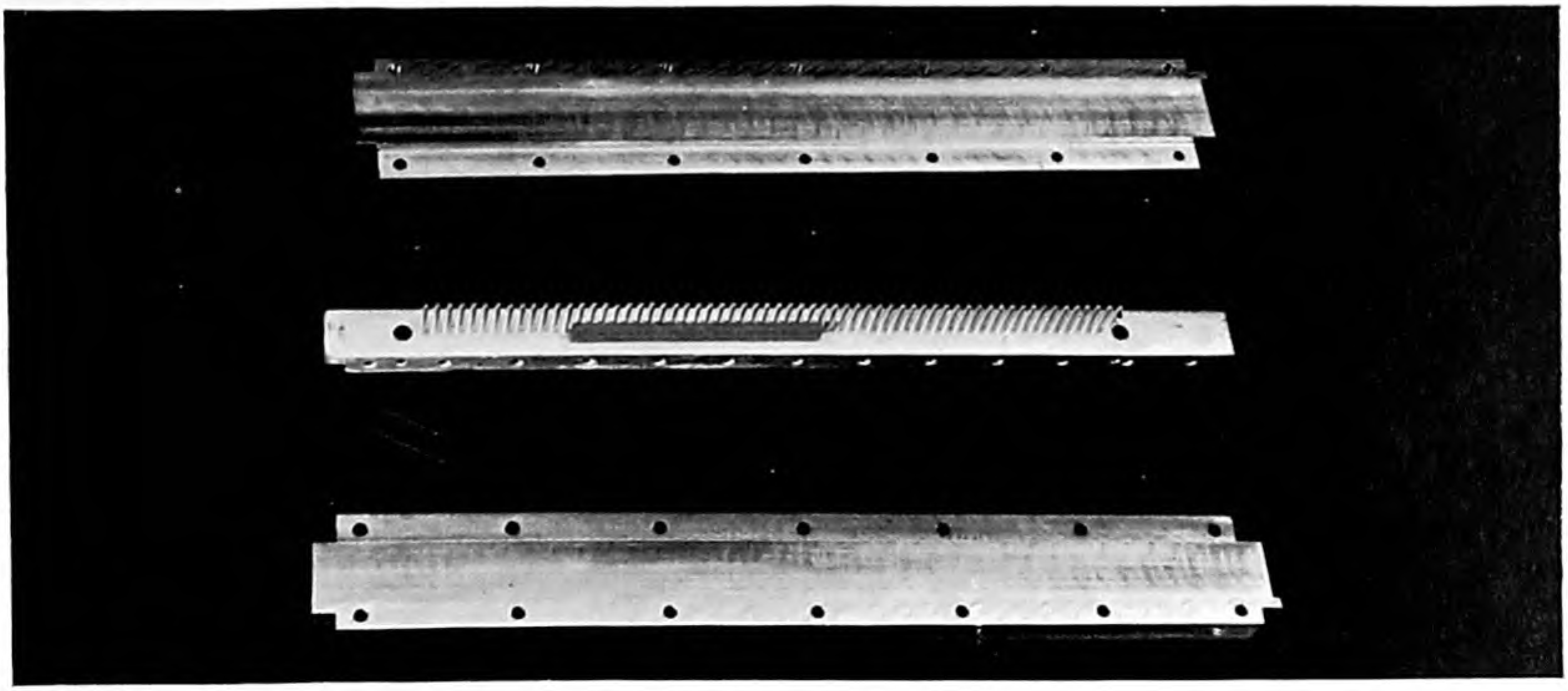


SOLID-STATE MASER consists essentially of a crystal *(center)* between the poles of a magnet *(top and bottom)*. Microwave energy of an appropriate frequency *(black curve at left)* pumps electrons in the crystal to a higher state. An input signal *(colored curve at left)* of lower frequency is amplified *(colored curve at right)* at the expense of the pumping energy *(black curve at right)*.

sponds to the energy gap between the top and middle levels will cause more downward than upward transitions, and the signal will be amplified. Pumping and amplification can go on at the same time, and so the maser operates continuously.

The first three-level maser was built at Bell Telephone Laboratories. Since then numerous masers of this kind, incorporating a variety of different crystals, have gone into operation at many other laboratories. One of them, as we have indicated, is already attached to a radio telescope. Soon they will be appearing in other applications.

There are jobs to be done by all the members of the maser family. In addition to simply telling time, ammonia and other gas-maser clocks will help explore some of the basic questions of physics. One plan is to recheck the celebrated Michelson-Morley experiment, which demonstrated that the speed of light is constant. Turning the maser's beam of molecules in two directions—along the path of the earth's travel and against it—



photograph is the maser cell, which is mounted at the right end of the waveguide. In the photograph at right the maser cell is dissected. One section of the large synthetic ruby of the maser stands against the row of pins in the middle; another one is to the left.

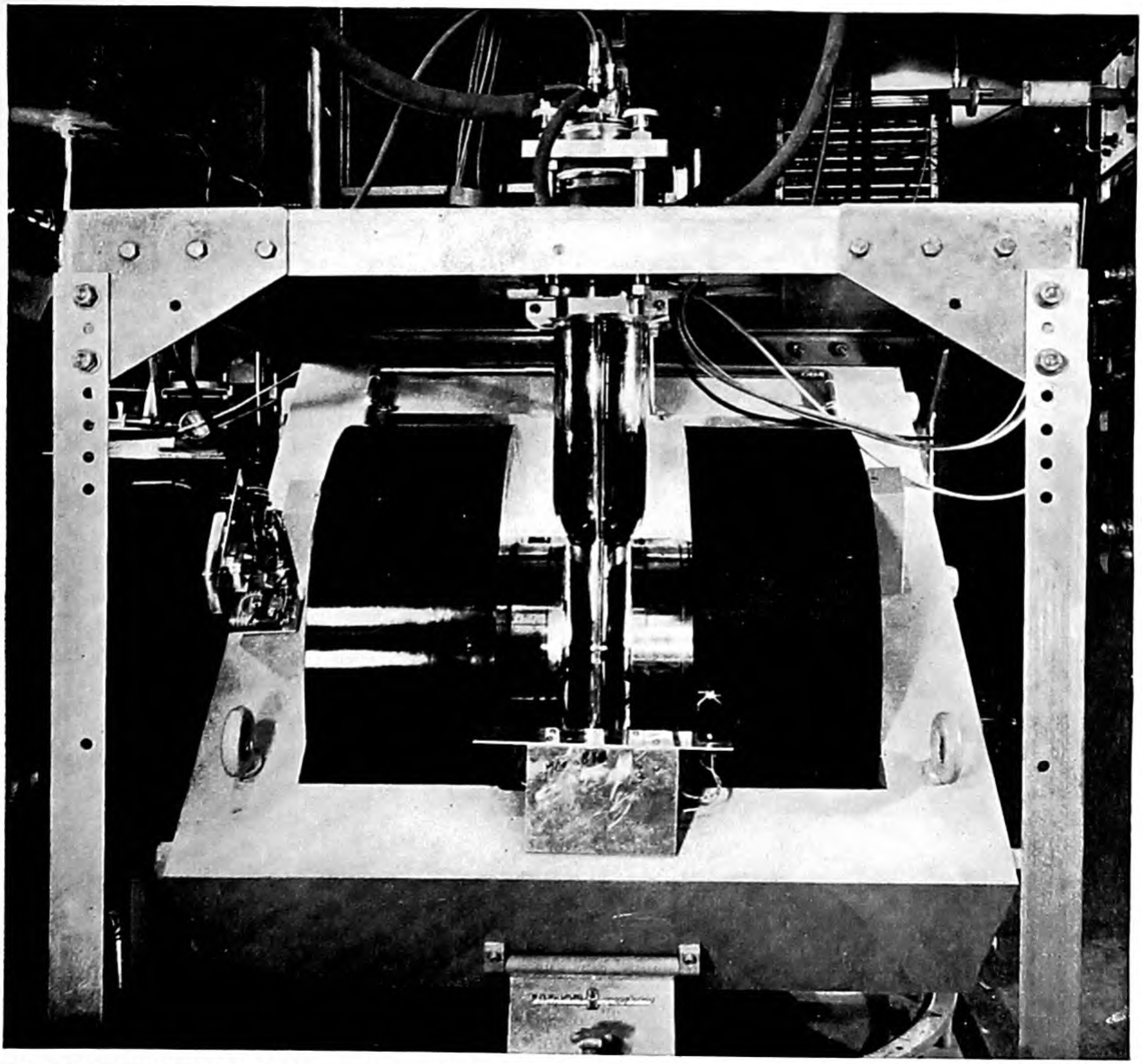
SOLID-STATE MASER is mounted at the focus of the 50-foot radio telescope of the Naval Research Laboratory in Washington, D.C. By largely eliminating the "noise" inherent in conventional amplifiers, the maser enables the telescope to detect very faint signals.

should result in no change of the output frequency, if light travels at a constant rate regardless of the motion of the observer. If there is a difference, it is too small to show up on Michelson's light interferometer. But the maser may be able to detect it. [As this issue of SCIENTIFIC AMERICAN went to press, it was announced that the experiment had been performed by Townes, working with J. P. Cedarholm, G. F. Bland and B. L. Havens of the International Business Machines Watson Laboratory at Columbia University. No difference was detected.]

Another project is a check on the general theory of relativity, which predicts that clocks are slowed up by gravitational fields. Artificial satellites will soon be circling the earth at distances where its gravity is noticeably weaker than at the surface. An atomic clock mounted in one of these vehicles could demonstrate the effect, if it exists.

Maser amplifiers may greatly simplify long-distance radio and television communication. As an example of what may be in the cards, suppose a ring of balloon satellites were made to circle endlessly around the earth. They would be permanent reflectors, from which signals could be bounced from any point on the earth's surface to any other. The received signals would be very weak. But the cold and lonely satellites would not contaminate them with much noise. Thus the sensitive, almost noiseless maser amplifiers could pick them up and boost them to useful levels without degrading them beyond recognition.

French workers have applied the maser principle to build a super-sensitive magnetometer for measuring the earth's field. In other laboratories people are thinking of using masers to produce beams of infrared radiation with an extremely narrow band of frequencies. The list is not exhaustive, and there are probably important applications that no one has thought of as yet. Quantum mechanics is adding a new dimension to "classical" electronics.



COMPLETE THREE-LEVEL MASER is photographed at Bell Telephone Laboratories. In center, between the poles of a large electromagnet, is a silvered flask which is filled with liquid helium. The maser cell is inside the flask between the two magnet poles.

The Author

JAMES P. GORDON was associated with C. H. Townes in the making of the first maser [see "Atomic Clocks," by Harold Lyons; SCIENTIFIC AMERICAN, February, 1957]. Gordon, then a graduate student under Townes at Columbia University, was especially concerned with investigating the properties of the ammonia molecule, which was the basic amplifying element of the early masers. Gordon comes from Scarsdale, N. Y., and is a graduate of the Massachusetts Institute of Technology. Upon receiving his Ph.D. from Columbia in 1955, he joined the staff of the Bell Telephone Laboratories, where he has worked on ammonia masers and on those questions of physics underlying the operation of solid-state masers.

Bibliography

ATOMIC CLOCKS. Harold Lyons in *Scientific American*, Vol. 196, No. 2, pages 71-85; February, 1957.

THE SOLID STATE MASER: A SUPER-COOLED AMPLIFIER. J. W. Meyer in *Electronics*, Vol. 31, No. 17, pages 66-71; April 25, 1958.

January 1959, Vol. 200, No. 1, pp. 46-53 OFFPRINT 216

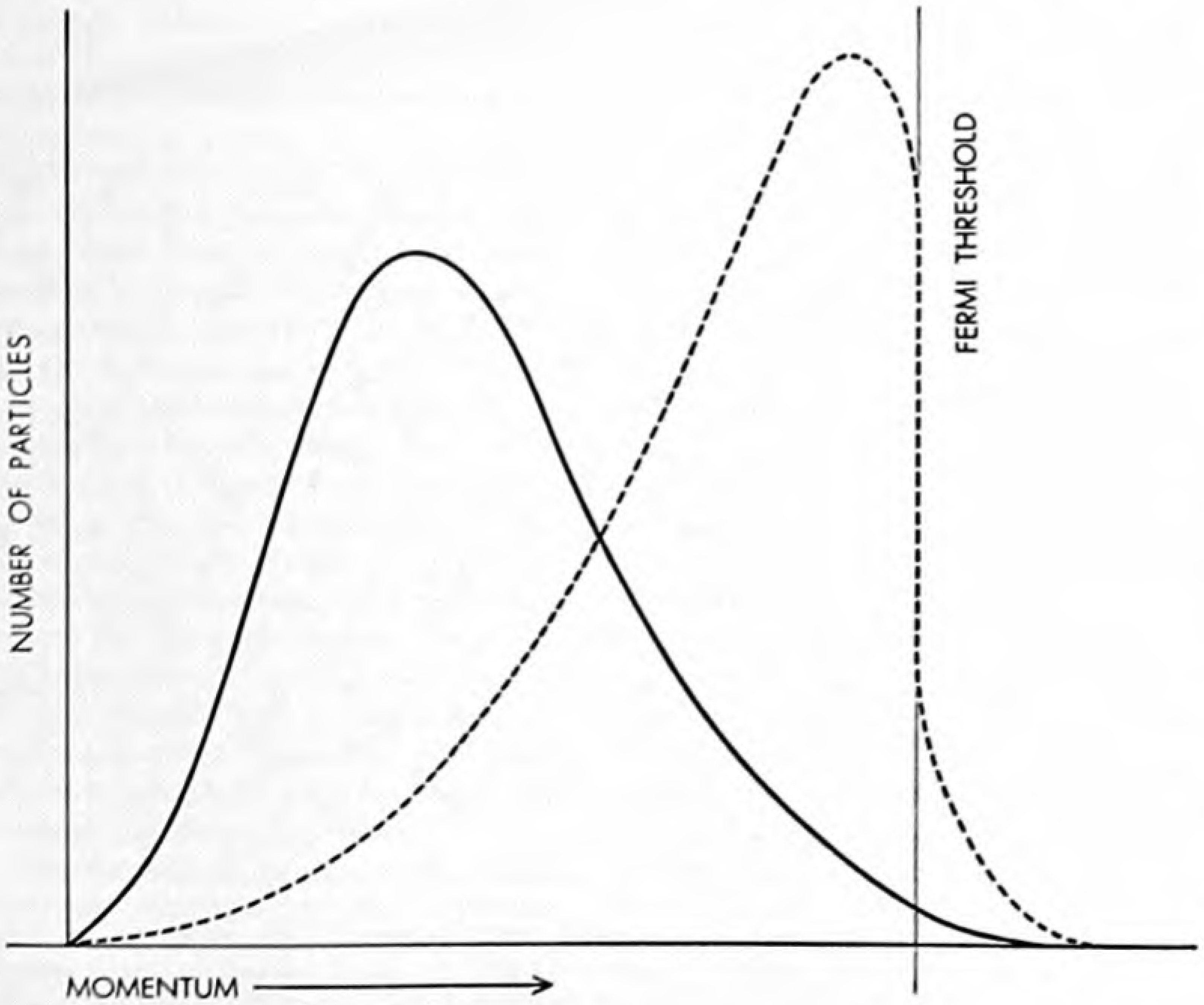
DYING STARS

by Jesse L. Greenstein

The white dwarfs, with their thermonuclear fires long since extinguished, are cooling down to the temperature of space. They are a portent of the inexorable fate of all other stars.

Here and there among the tens of thousands of stars in the nearby regions of our galaxy are a few hundred whose fires have gone out. Once they burned as brilliantly as any we now see in the sky. Some had the "normal" size and brightness of the sun; some were giants, with many times the sun's diameter and brightness. Now these stars are approaching the end of the road. They have exhausted their fuel. The inward pull of gravity, no longer opposed by the outward push of pressure generated by heat within, has shrunk their diameters to a tiny fraction of stellar size, to that of the earth and even smaller, compressing their huge masses to unimaginable densities of many tons per cubic inch. In their fading light, detectable only by the instruments and techniques of modern astronomy, they are radiating the heat still left from the past out into the cold reaches of space

We call these stars "white dwarfs." They hold clues to many interesting questions of astrophysics. Until recently, however, much of what we "knew" about them was the fruit of theoretical speculation. They comprise some 3 per cent of all the stars in our galaxy and so must be rated a common type. Yet their luminosities are so low that only a few hundred have been tentatively identified and only 80 observed in any detail. Study of their color and the lines detectable in their spectra is yielding new insight into the synthesis of elements in younger stars. Their densities represent states of matter which we can hardly think of duplicating in terrestrial laboratories. But the white dwarfs have a more general significance. They are a portent. They show us that the laws of thermodynamics, which circumscribe events on the minuscule scale of our planet, hold also as the inexorable plan of the life history of the stars.



MOMENTUM OF PARTICLES in a "perfect" gas (*solid line*) follows the bell curve of random distribution. In a "degenerate" gas (*broken line*), the curve shows fewer low-momentum states available. Only the few particles above the Fermi threshold move at random.

An irreverent physicist once rephrased the laws of thermodynamics to read: (1) you can't win, (2) you can't even break even, (3) things are going to get worse before they get better and (4) who says things are going to get better?

When it is applied to stellar processes, the first law reminds us that stars do not create energy, but only convert energy from one form to an equivalent quantity of another form; that is, they convert to radiant energy the energy contained in their gravitational potential and in that fraction of their mass which is consumed in thermonuclear reactions. They can never produce more energy than they start out with. In a steady-state star, with a stable balance between its gravitational contraction and the pressure generated by the heat within, the expenditure of thermonuclear energy can go on for a long time—10 billion years in the case of the sun.

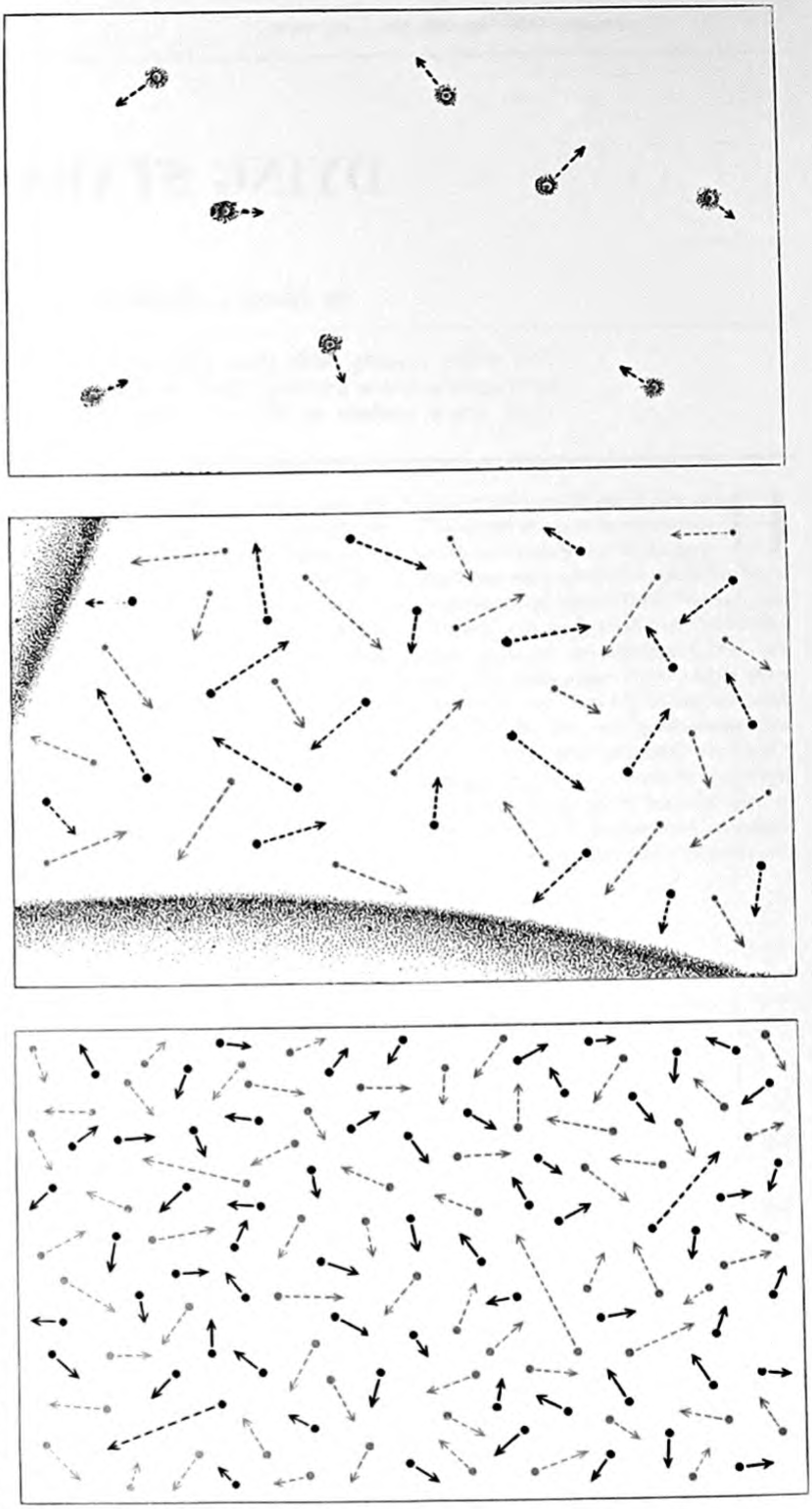
But the second law reminds us that this cannot go on forever. A star can never recapture the energy it wastes into the sink of space; its life history is irreversible. As it uses up the hydrogen that comprises the bulk of its substance,

the thermonuclear furnace begins to falter. Gravitational contraction restores the equilibrium, converting potential energy into thermal energy. But contraction raises the density of the star, and the new balance between gas pressure, heat transfer, energy production and radiation loss changes the internal structure. The star brightens, its outer envelope grows larger, and stellar "evolution" begins—earlier in the life of brighter stars, later in that of the fainter ones.

As the star enters the last phase of its existence it shrinks to the final, stable configuration of a white dwarf. The third and fourth laws of thermodynamics now assume increasing relevance to its condition. The third law says that the star will ultimately cool down to the temperature of space, and the fourth law declares that it will then no longer give forth light or heat. At this terminal point the white dwarf becomes a black dwarf. Since we could not observe black dwarfs, if there are any, we shall not now give further consideration to them. In any case a star persists as a white dwarf for billions of years. Its structure and condition in this phase is what interests us here.

Matter at white-dwarf density is strange to contemplate by celestial as well as terrestrial standards. A star like the sun has an average density of almost one gram per cubic centimeter, about the same as that of water. Astrophysicists nonetheless find it feasible to deal with the behavior of solar matter as if it were a gas, with its particles free to move about at random. At the high temperatures of the solar interior, hydrogen is 97 per cent ionized; the electrons of nearly all the hydrogen atoms are stripped from their nuclei (protons). This means that the bulky structure of the hydrogen atom, 10,000 times the diameter of its constituent particles, is obliterated. As a result a cubic centimeter of ordinary stellar material is largely empty space. The tiny protons and electrons are free to move in all directions and at all velocities, just as they would in a highly rarefied gas.

In a white dwarf, on the other hand, a mass on the order of the sun, equal to 332,000 earth masses, may be packed into a volume no larger than that of the earth, which has but one millionth the sun's volume. The density ascends to 1,000 kilograms per c.c.—more than 15 tons per cubic inch. Even after a white dwarf has cooled below the temperature needed for ionization, the atoms remain dissociated under the crushing pressure of gravity. The particles are not yet so tightly packed, however, that their vol-

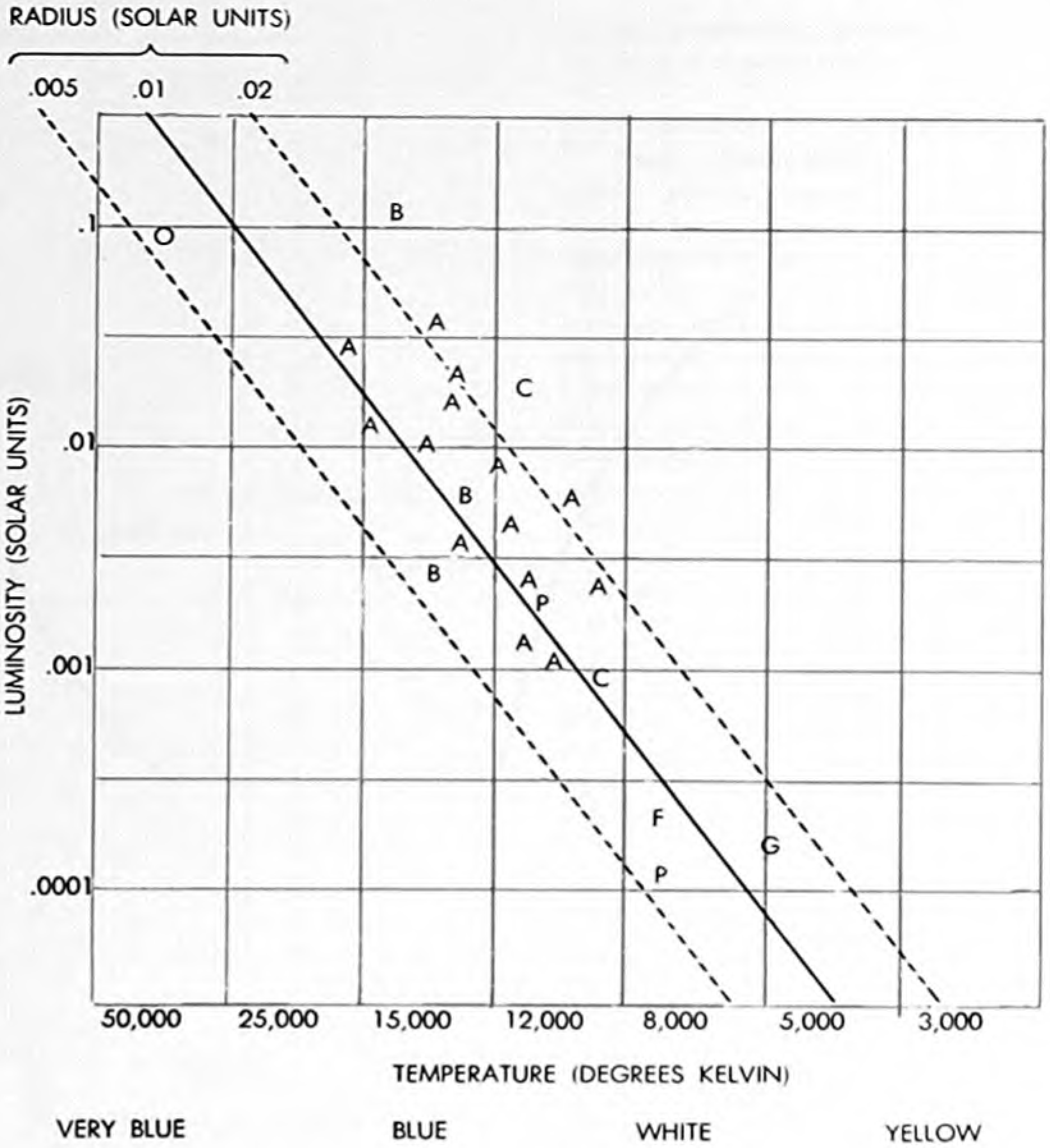


"DEGENERATE" GAS (*bottom*) is contrasted with "perfect" gases, made up of atoms (*top*) and ionized particles (*center*). Space available in gas of normal pressure permits random motion (*broken arrows*) to atoms. In an ionized gas, even at the density of a solid, the obliteration of the structures (*shadowed arcs*) of all but a few atoms opens up space to permit random motion of electrons (*black*) and nuclear particles (*color*). At the extreme density of a degenerate gas the energy states of most electrons are prescribed at low momenta (*solid arrows*). Only the nuclear particles and a few electrons move at random.

umes overlap; there is still empty space between them. But because each particle has only a small volume of space in which to move, its momentum as well as its position is prescribed. The exclusion principle of physics, which rules that no two particles can occupy the same energy state, rigidly specifies the coordinates and motion for all low-momentum states Since the electrons are the lighter particles, they have the lowest momenta and are frozen in space and velocity. Collisions cannot result in arbitrary changes of momentum, but can only kick the electrons into unoccupied states. A few electrons which attain velocities approaching that of light, above the so-called Fermi threshold, are still free to move, as are the nuclear particles [*see illustration on preceding page*]. The gas has entered the "degenerate" state.

We owe to Subrahmanyan Chandrasekhar of the Yerkes Observatory a beautifully complete theory of a self-gravitating degenerate sphere of gas. Strangely, according to the theory, the greater the mass of a white dwarf, the smaller its radius. This follows, however, from the degenerate-gas law, which predicts a gas pressure, for a given density, sufficient to counteract gravitational pressure only when the star is greatly collapsed. The inverse relationship of mass to radius is not affected, as it is in other stars, by temperature, luminosity or energy production. The mass and hence the radius of a white dwarf is fixed, in the theory, by the elemental composition of the star. For stars of each composition there is an upper limit of mass. Calculation from the theory shows, for example, that a white dwarf composed of hydrogen would have a maximum possible mass 5.5 times that of the sun. On the other hand, a white dwarf made up of heavier elements should have no more than one fourth this mass, or 1.4 solar masses. A more massive star must lose mass or suffer a catastrophe before it becomes a white dwarf. We have few reliable determinations of white-dwarf masses, but all such determinations lie well below the theoretical maximum of 1.4 solar masses. This is important confirmation for the deduction that these stars have exhausted their hydrogen, the principal thermonuclear fuel.

The theoretical picture of the white-dwarf star, extended by other investigators, makes it clear that it will always be difficult to test theory by observation. The dense degenerate mass of the star is surrounded by a sharply differentiated envelope about 65 miles deep; the material here is nondegenerate because of the

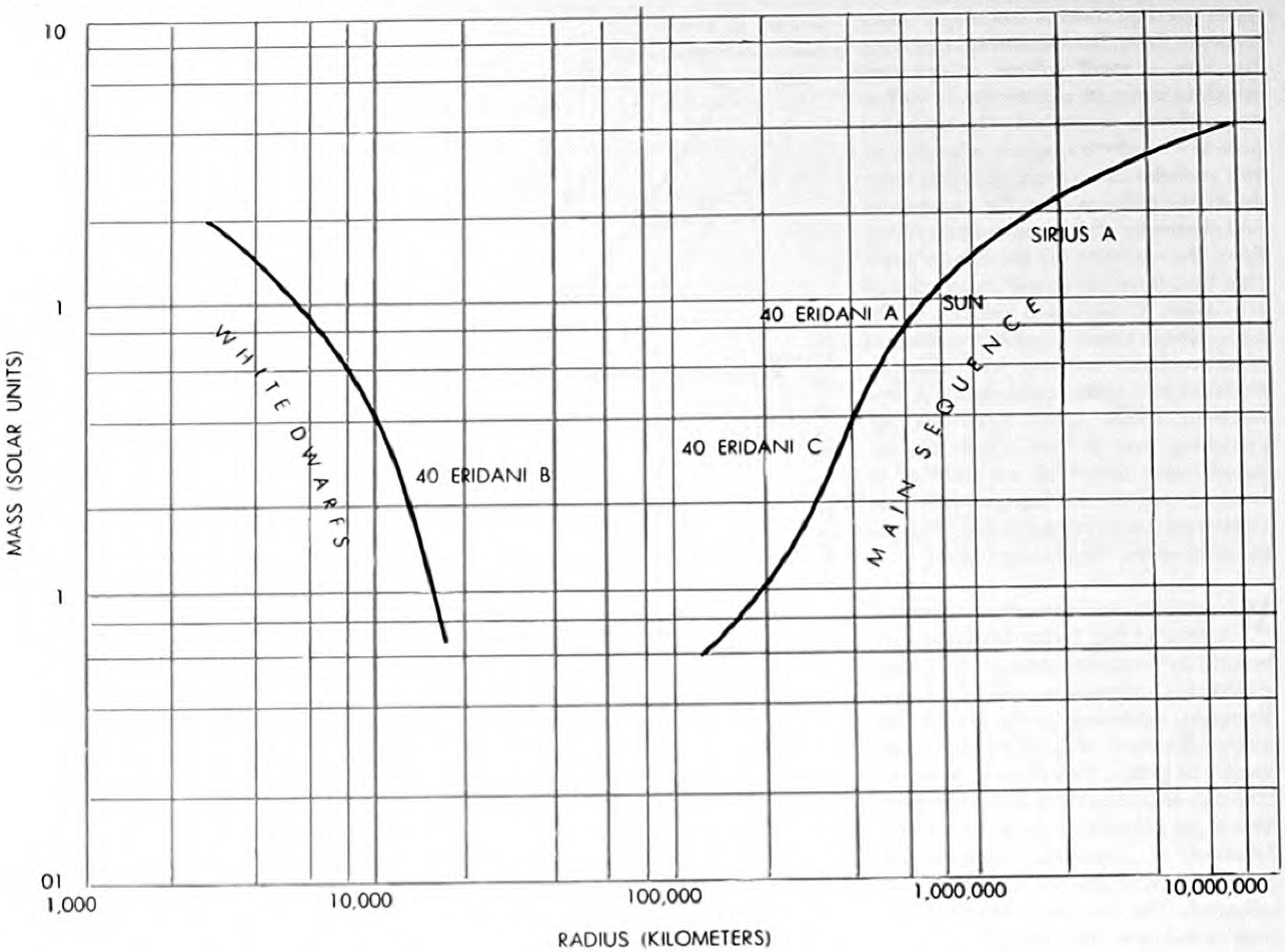


RADIUS AND TEMPERATURE of white dwarfs show no correlation. Stars of various radii occur at all temperatures as indicated by the positions of the letters standing for various types. This is evidence that dwarf stars cool down without further gravitational contraction.

lower pressure. Superposed on the envelope is the atmosphere of the star, which is only a few hundred feet deep. This is the only part of the star we can study spectrographically. What we observe in the spectra of normal stellar atmospheres, which are thousands of miles deep, tells us much about their surface temperature and composition, and also a good deal about their interior. The shrunken atmosphere of a white dwarf bears small relevance to the interior and can tell us little about it.

Evry Schatzman of the Institut d'Astrophysique in Paris has shown that white dwarfs cannot have the same composition at their surface as in their interior. In the absence of convection the gas stratifies under the intense gravitational field. The residual hydrogen is squeezed to the surface, while the helium and heavier elements gravitate to the center. Were it not for electrical forces, the electrons would tend to float on top. The electrical fields and nuclear forces set up by the stratification contract the star still further and so reduce the maximum possible mass to 1.25 solar masses.

The fading light that carries off the heat remaining in their interiors has given us the location of several hundred possible white dwarfs. The brightest of them has a luminosity only .01 that of the sun; the faintest known dwarf has only .0001 solar luminosity, so faint that such stars cannot be observed at distances greater than 30 light-years. Their low luminosity, combined with our theoretical knowledge of their internal structure, provides convincing evidence that they have ceased transforming matter into energy. At their high densities thermonuclear reactions would go on at enormously high rates, even if temperatures were as low as 10 to 30 million degrees Kelvin. The reaction rate would be even further increased by the dense packing of the electrons, whose negative charges would partially nullify the mutual repulsion of the nuclei. The only possible explanation of their low lumi-



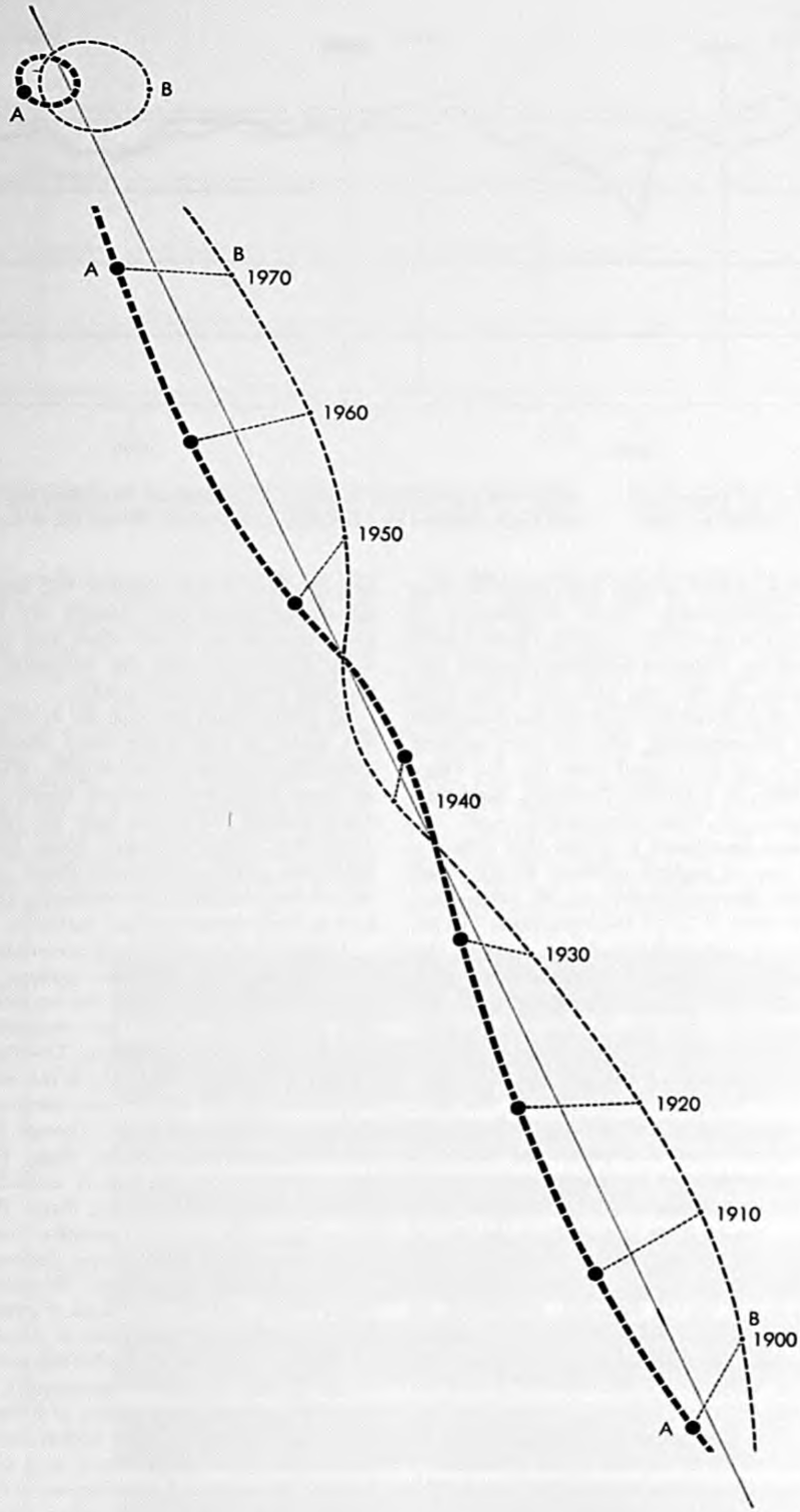
MASS AND RADIUS of white dwarfs show a correlation exactly opposite to that of normal "main sequence" stars (*curve at right*). The latter show increase of radius with increase of mass. White dwarfs, in contrast, have smaller radii at higher mass. The smallest dwarfs have masses which are larger than that of the sun, but these masses are compressed into volumes smaller than that of the earth.

nosity is that hydrogen must now comprise less than .00001 of the mass of a dwarf star. Reactions involving heavier elements—such as carbon, oxygen, nitrogen and neon—require higher temperatures than are likely to occur, though helium might react with these in large concentration at very high densities. However, another set of theoretical considerations argues against the possibility of any energy production at all. In a normal star the thermonuclear reaction-rate is regulated by feedback; with increase in temperature the star expands, and the reaction rate is damped. In a degenerate gas, on the other hand, pressure is unaffected by temperature. Local heating would bring higher temperature and an increase in the reaction rate. The star, in consequence, would explode. We must therefore conclude that the white dwarfs have substantially exhausted their nuclear-energy sources.

Because their luminosity is so low, it is difficult to obtain detailed information about other aspects of the dwarf stars from spectrographic analysis of their light. Only about 80 such stars have been studied in detail. With the light-gathering power of the 200-inch Hale telescope on Palomar Mountain I have observed 50 white-dwarf spectra at a larger scale than any obtained before.

Spectrographic analysis establishes with certainty that the white dwarfs are dwarfs indeed. The derivation of radius from the spectra is somewhat indirect, but it is reliable. Both from photoelectric analysis of the color of the light and study of the behavior of the absorption lines we can determine temperature. From apparent brightness and from independent measurement of distance, we establish the true luminosity. By combining temperature and luminosity, we determine radius. The results are impressively monotonous: the well-determined radii all lie between 3,000 and 10,000 miles. The constancy of dimension is in contrast to the range of size in normal stars, from .1 to 10 times the radius of the sun (430,000 miles) for "main sequence" stars [*see illustration on this page*], and on up to 10,000 times for red giants. The smallest white dwarf known has an estimated radius of only 2,800 miles, much smaller than the radius of the earth. This is close to the theoretical minimum for a star that has exhausted its hydrogen; the radius indicates a mass of 1.2 solar masses and a central density of 150 tons per cubic inch.

One of the most important theoretical predictions is fulfilled with the finding that there is no dependence of radius on surface temperature. The dwarfs we have observed range in temperature from 50,000 to 4,000 degrees K. The hottest is a blue-white star in the earliest phase of white-dwarf evolution; the coolest, a faint, reddish-white dwarf. As plotted in the illustration on the opposite page, stars of the same radius appear down the full range of temperature. Since their initial masses may vary, it is clear that they start with a small spread of radii at the upper left corner of the chart and cool off without further gravi-



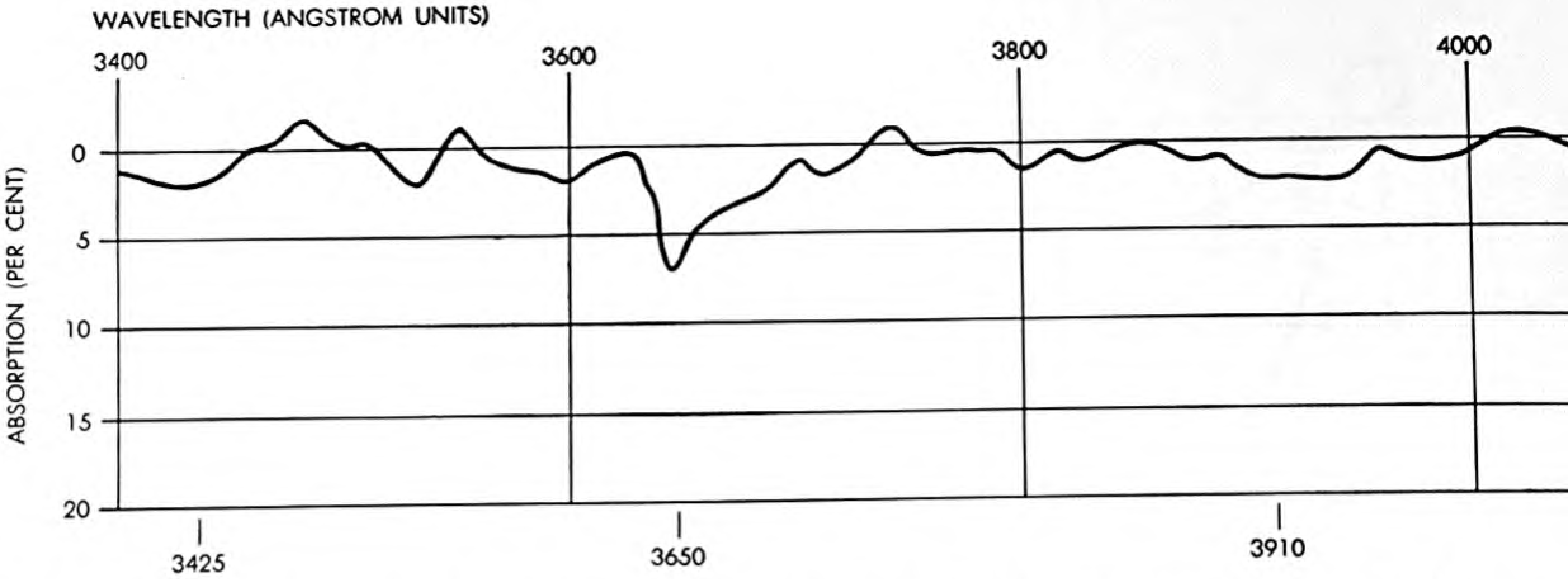
DOUBLE-STAR SYSTEM of Sirius is composed of one of the brightest stars in the sky (A) and a white-dwarf companion (B). Their orbits around the center of gravity of their system is shown at top. The motion of the two stars and of the center of gravity of their system with respect to the earth is indicated by the broken lines running diagonally up this diagram.

tational contraction downward and to the right in straight lines.

Unfortunately it is impossible to match these measurements of radius to equally reliable observational determinations of mass. Newton's laws can give the masses from observed orbital motion only in the case of those stars that are members of multiple systems. Three such dwarfs are known. For two of them, Sirius B and Procyon B, the masses are reliably established at 1 and .65 solar mass respectively. But their major companions, Sirius A and Procyon A, are so bright and so close that the spectrographic plate cannot register an uncontaminated picture of either of these two dwarfs. As a result it is still impossible to measure their radii.

The best-known white-dwarf member of a multiple system belongs to a three-star group: 40 Eridani. Here, fortunately, the distances between stars are wide enough so that good spectra can be obtained, and yet close enough for orbital motion to give reliable measurements of mass. From analysis of the spectrum, I have derived a radius of 6,500 miles, .016 of the solar radius; gravitational measurements establish the mass at .45 solar mass. Calculation from the theoretical mass-radius relationship yields a mass of .39 solar mass, satisfactorily close to observation. Thus, at least in the case of the single star that permits complete test by observation, the well-articulated theory of white dwarfs finds solid support.

The spectra of the white dwarfs also confirm in a general way the theoretical prediction of their elemental composition. One type either shows no hydrogen lines at all, or has hydrogen lines which indicate the presence of relatively tiny residual quantities of hydrogen. Compared to the spectra of normal stars, in which hydrogen lines are universally strong, this anomaly would be enough to identify the dwarfs as a genus apart. The spectra of the commonest type of white dwarf (Type A), however, show only the residual hydrogen and no heavy elements. Here, apparently, gravitational forces have pulled all of the heavier elements, even helium, out of the atmosphere and squeezed the hydrogen to the surface. In dwarfs with surface temperatures below 8,000 degrees, the hydrogen lines vanish completely, and we see only a few lines due to metallic elements. Ross 640 is such a star [*see illustration on page 131*]; it is hot enough to show hydrogen lines if any hydrogen were present. In general the spectra of white



ANOMALOUS SPECTRUM of a white dwarf shows no absorption lines, but does show diffuse bands of absorption at points not associated with any familiar elements or compounds. The spectrum has here been analyzed by a sensitive photoelectric device which meas-

dwarfs reflect little of the regular correlation between line characteristics and temperature found in normal stars. The varied compositions of their atmospheres therefore may be taken as evidence of their evolutionary history. From the spectrum of Ross 640 we can deduce that this star and other stars like it turned to synthesizing heavy elements from helium after exhausting their hydrogen. The redder and still fainter star called van Maanen 2 (VMa2) is the coolest so far subjected to detailed spectrographic analysis. Its peculiar spectrum [*see illustration on opposite page*] indicates that this star began as a metal-poor member of the long-lived, stable Population II family. Since its present low luminosity gives this star an age of four billion years in the white-dwarf phase alone, van Maanen 2 must have lived out its entire life as a brilliant star before the sun and the earth were formed. In a still fainter, cooler and more ancient star, no lines have yet been detected with certainty.

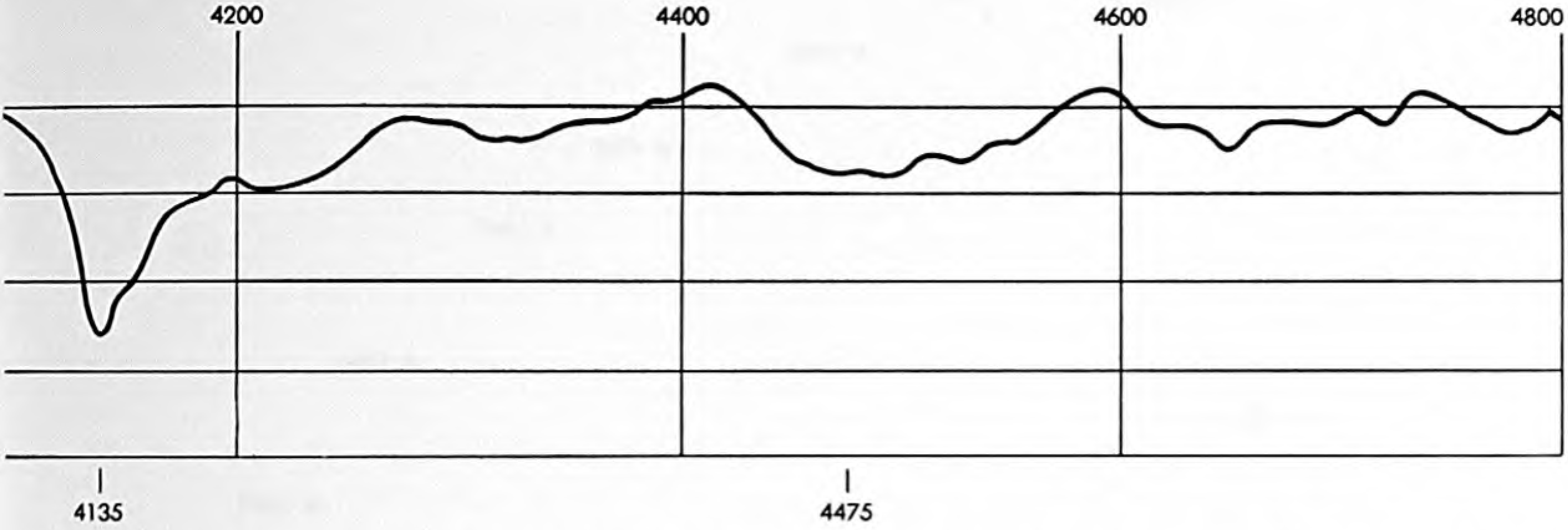
A spectrum without absorption lines might seem to be of academic interest to astrophysicists, who employ these lines as the tools of their trade. But we have spent many nights observing and many months of analysis to establish the real absence of lines in six white-dwarf spectra. Subjected to the most sensitive photoelectric inspection yet possible, the plates show no line, band or absorption depression as deep as 5 per cent. There are a number of possible explanations. Perhaps the most satisfactory will be found upon closer inspection of lines that do appear in other white-dwarf spectra. The extreme broadening and attenuation of the hydrogen lines in some spectra helps to make the complete disappearance of lines at very high pressure more understandable. Such broadening of lines is caused by random electric fields and by collisions between charged particles. In the van Maanen 2 spectrum Volker Weidemann of the Bundesanstalt in Braunschweig, who has been working with us on a grant from the Air Force Office of Scientific Research, has found lines of iron, magnesium and calcium broadened in a way that indicates a rate of particle collision 10,000 times that observed in the sun. He estimates a pressure of 2,000 atmospheres in this peculiar atmosphere—dense enough for some molecules to form. But though metal lines may be thus broadened, it is surprising that they should disappear entirely, as they do in the six spectra that show no lines at all.

To compound the mystery we have come upon several spectra with diffuse, shallow bands that cannot be related to any established laboratory spectral line; the photoelectric tracing of a plate made for one of these is shown at the top of these two pages. These bands may originate from molecules or unstable free radicals under unusual conditions of temperature and pressure. How atoms behave in the strange environment of the white-dwarf atmosphere is not yet known.

Our generation has seen at least one star arrive at the end of the evolutionary road and become a white dwarf. The recurrent nova, WZ Sagittae, which exploded in 1913, exploded again in 1946, brightening about 1,000 times. Its brightness is now about .01 that of the sun, and its spectrum resembles that of white dwarfs in everything but the presence of superposed emission lines. These lines are presumably due to the continued ejection of hot material. WZ Sagittae demonstrates one, though not the only, process by which stars may lose their mass and make the transition to the final stage in their history.

As living things live and die in countless ways, so stars have many possible evolutionary histories and deaths. When we have learned to read the spectra of white dwarfs better, we may see what paths they have traveled. Their faint light may give us evidence which will show what processes went on during ages past in their thermonuclear furnaces.

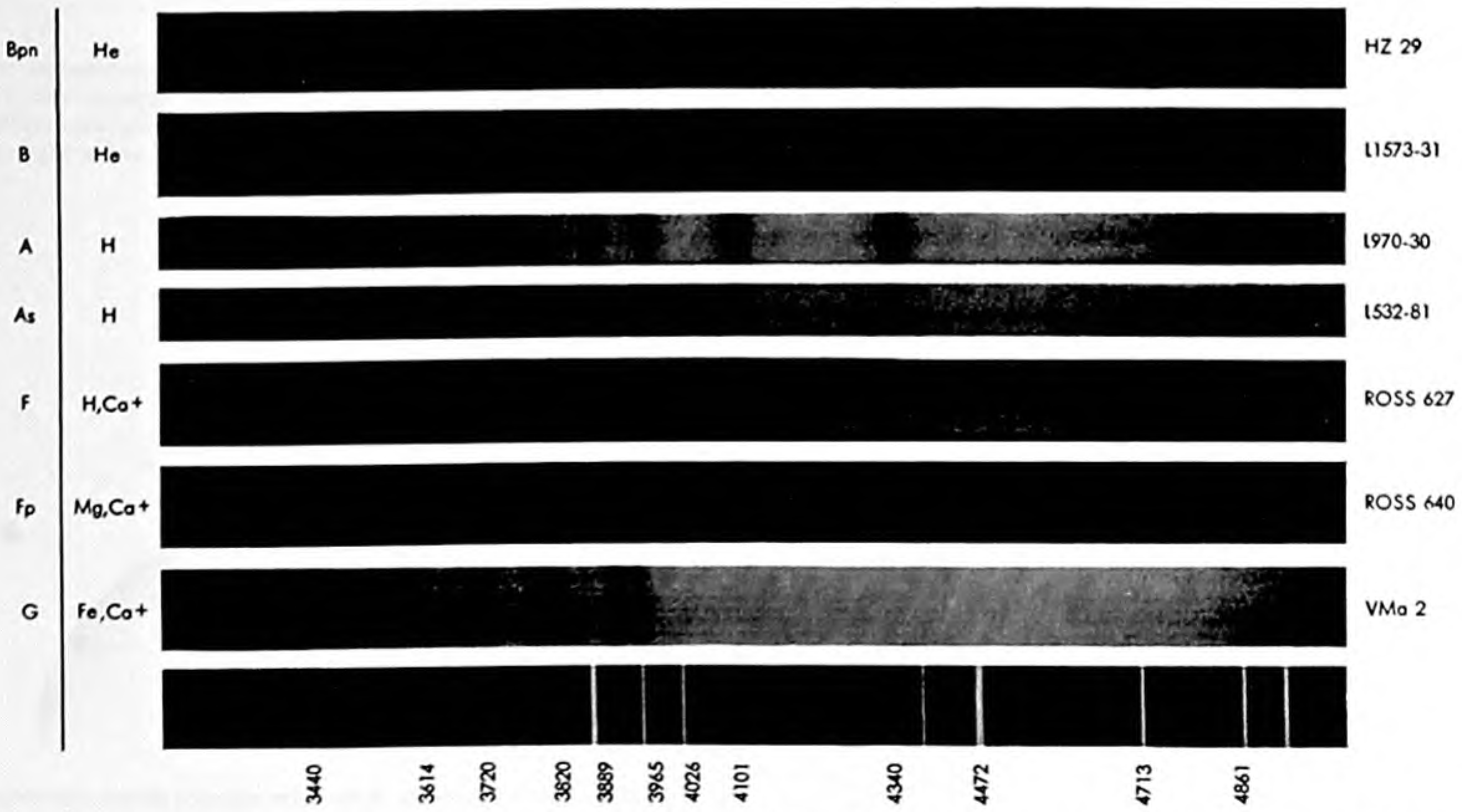
A white dwarf takes a long time dying. Its light bespeaks the slow leakage of heat from its interior down the temperature gradient set up by the conductive opacity of the degenerate gas. The thermal energy is contained only in the nondegenerate nuclei and the few electrons above the Fermi threshold. Though the initial temperature may be high, this thermal energy is all that is available throughout the entire dying stage. But as the star cools and its luminosity fades, the temperature gradient also declines. The dissipation of energy therewith slows down, and the time scale of evolution toward lower luminosity is greatly extended. According to Martin Schwarzschild of the Princeton Observatory, a white dwarf composed mainly of helium takes three billion years to cool from its initial blue-white stage down to a surface temperature of 7,000 degrees in the yellow-white stage. From yellow down to the 4,000 degrees of the faintest known red-white dwarf, it takes another five billion years. But 4,000 degrees is still red-hot. From red to infrared, the star will fade over fantastic spans of time, large compared to any present estimate of the age of our galaxy.



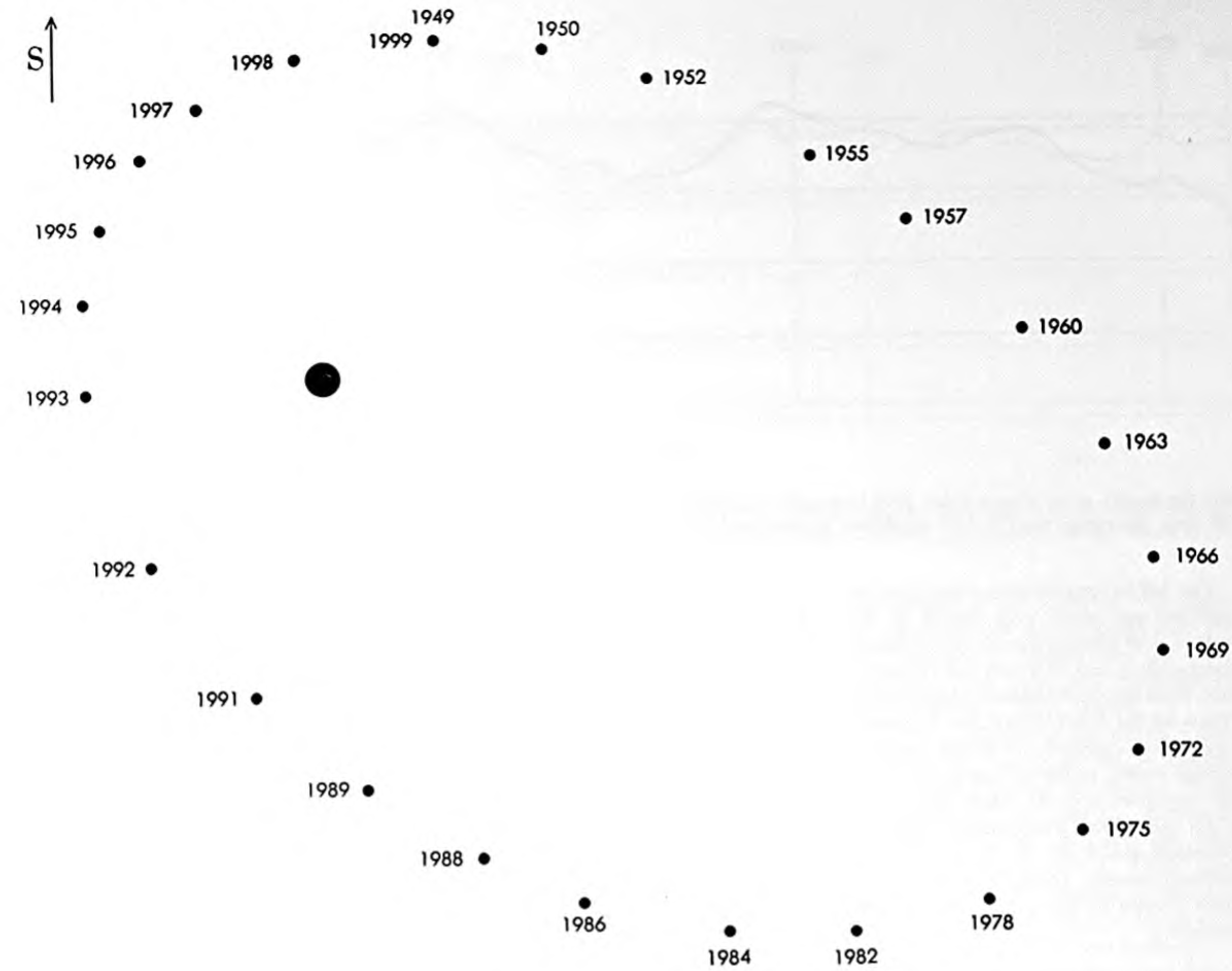
ures the density of the photographic plate from point to point. Only the deep absorption band at 4135 angstroms wavelength is visible to the eye on the spectrographic plate. The absorption bands may be due to presence of elements or free radicals under high pressure.

The fall in temperature brings the degenerate gas phase ever nearer to the surface. The nondegenerate electrons become scarcer and, at a very low temperature, even the nuclei become degenerate. When all the nuclear particles and electrons have occupied the lowest possible energy states, radiation ceases and the star becomes a giant "molecule." This is the end of the irreversible process of evolution—proof of the fourth law of thermodynamics. There are, however, no black dwarfs in our galaxy; it is as yet too young.

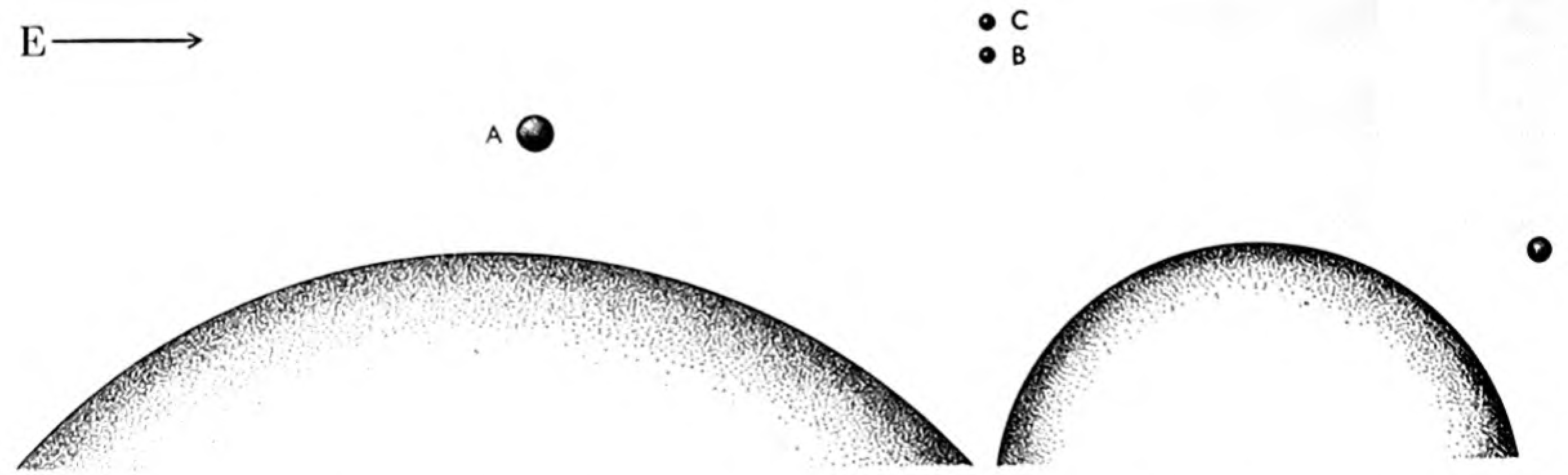
On the one-way track described here, all stars eventually fade to extinction. How will the sky look after our sun's evolution is complete, and our dead planets circulate about a dying star? In about seven billion years the sun will be a hot and very blue-white dwarf, too small to show a disk to the unaided eye on earth. The earth's temperature will be about 300 degrees below zero Fahrenheit. The sky at night will no longer be filled with stars, since star formation will have ended, and the high-luminosity stars that comprise our constellations will long ago have disappeared. Probably no star will be visible, except for an occasional faint, red normal main-sequence star that passes by chance near our dying system; such stars are so faint that their nuclear energy suffices for thousands of billions of years. Although the formerly bright stars will have become white dwarfs, they will all be too faint to be seen, and black night will reign supreme. Yet close to one of the faint red stars life might exist on other planets, in forms and for ages unimaginable to us.



DWARF STAR SPECTRA **of various types (identified by initials at far left) show absorption lines for only a few elements (identified by initials in second column at left). The individual stars are identified by their code numbers at right. The spectrum at bottom is the reference spectrum of helium and hydrogen. Absorption lines of Type A dwarfs are characteristically diffuse and broadened.**



DWARF STAR IN SIRIUS traces the orbit shown here with respect to the large primary star of this double-star system. The dates give the location of the dwarf in its orbit through the second half of this century. Its close approach to the primary star in recent years has made it impossible to secure spectrographic images uncontaminated by the light flooding from the 100-times brighter primary star. As the dwarf star approaches the apogee of its orbit during the next 20 years, it may be possible for astronomers to secure better spectra.



TRIPLE-STAR SYSTEM in constellation Eridanus is composed of a bright primary normal star (A), a faint late-type star (C) and a white dwarf (B), which appear in the relative positions, but not to the scale, indicated by the small spheres at the top of this diagram. The relative diameters of the three stars are shown across the bottom of diagram, star A (*at left*) having a radius .9 that of our sun; star C (*second from left*) having a radius .4 that of the sun; and the dwarf having a radius .017 that of the sun, or 7,000 miles.

The Author

JESSE L. GREENSTEIN is an astronomer at the Mount Wilson and Palomar Observatories and, in addition, heads the astronomy department at the California Institute of Technology. A New Yorker, he graduated from Harvard College in 1929, took an M.A. at Harvard in 1930, then rode out four depression years as an operator in real estate and investments. In 1934 he returned to Harvard for his Ph.D. Subsequently he joined the staff of the University of Chicago's Yerkes Observatory, first as a National Research Fellow, then as associate professor. Greenstein has worked at Mount Wilson and Palomar and Cal Tech since 1949; for the past six years he has chaired the International Astronomical Union's commission on stellar spectra.

Bibliography

AN INTRODUCTION TO THE STUDY OF STELLAR STRUCTURE. S. Chandrasekhar. The University of Chicago Press, 1939.

THE SPECTRA OF THE WHITE DWARFS. Jesse L. Greenstein in *Encyclopedia of Physics*. Springer-Verlag, 1958.

WHITE DWARFS. Evry Schatzman. Interscience, 1958.

SCIENTIFIC AMERICAN July 1956, Vol. 195, No. 1, pp. 55-68

THE ATOMIC NUCLEUS

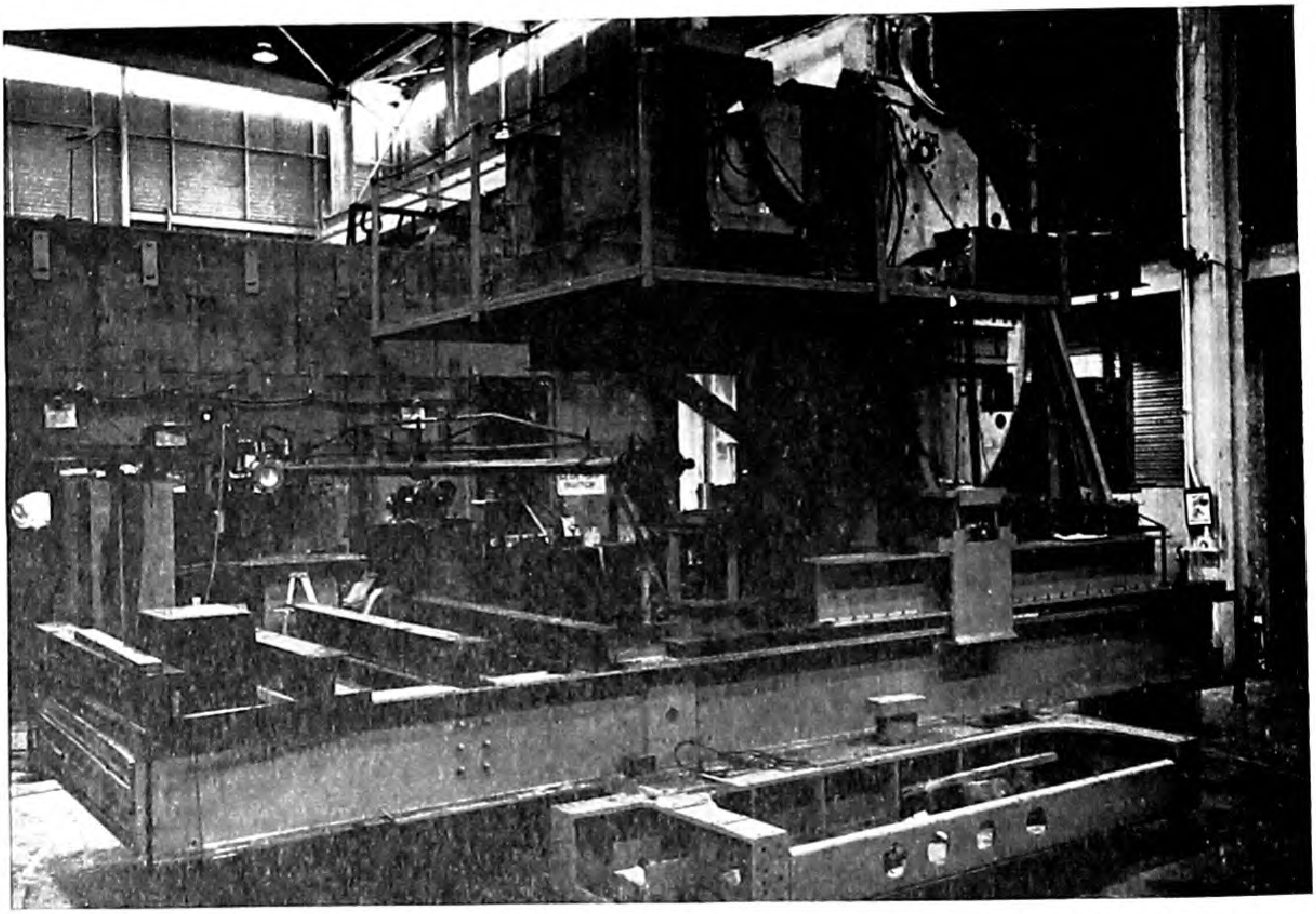
by Robert Hofstadter

Its structure, when examined by a beam of high-energy electrons, is characterized by a fuzzy "skin," the density of which decreases from the inside out. Even individual protons have this construction.

Not much more than 50 years ago it was still possible for leading physicists and chemists to argue over whether atoms really exist. Today the most backward schoolboy knows that atoms are real. He even knows what they look like. The picture of a little round nucleus surrounded by a cloud of electrons is practically the trademark of our time.

It is not just a popular emblem. Physicists also have a mental image of the atom which is much like this one. In fact, they have gone farther. For the past 20 years or so they have been trying to draw mental pictures of the inside of the nucleus.

That physicists can even dream of perceiving details in so minute an object is a tribute both to their imagination and to the delicacy of their experimental



MAGNETIC SPECTROMETER measures the scattering of high-speed electrons from target nuclei. Electrons from the Stanford University linear accelerator enter through the thin pipe which extends from the left background to the center foreground. The target material is suspended in the ladder-like holder in front of the pipe. The huge, D-shaped magnet which focuses the scattered electrons can be seen at the far right. It is carried on a naval gun mount which moves it to various angular positions around the target.

technique. The smallness of the nucleus has been pointed out many times, but it is always worth emphasizing when one is trying to appreciate what nuclear physics is about. The diameter of the nucleus is a few ten-trillionths of a centimeter. If the nuclei of all the atoms in the earth could be stripped of electrons and packed together, they would make a ball only 200 feet in radius. We usually think of an atom as a very small object indeed. Yet if an atom could be expanded so that its outer electrons enclosed an area the size of New York City, the nucleus at its center would be about as big as a baseball.

That it is possible to peer within this speck of matter is one of the most impressive feats of modern physics. It underlines the genius of Lord Rutherford and other early investigators, who accomplished the feat with the relatively crude methods which were available to them. Recently the author and his colleagues at Stanford University, using the advanced technology of present-day experimental physics, have developed a new and very powerful instrument for examining nuclei. With it we are getting a look at details that have never been seen before, and which show that older pictures of the nucleus must be revised. We have even begun to penetrate the interiors of the "ultimate" particles—the protons and neutrons of which nuclei are made!

Models of the Nucleus

The significance of our work is best understood against a background of ideas about the nucleus that have been developing since the early 1930s. In the first place, it should be said that terms like "looking" into the nucleus or forming "pictures" of it are pure metaphor. The nucleus is utterly and hopelessly invisible. In fact, the physicist does not speak of "pictures" but of "models." This word is better because it reflects the indirect approach he is obliged to take. His experiments, as we shall see, do not yield a direct representation of the nucleus. The physicist must consider separate sets of experimental results and then try to imagine a model of the nucleus that would account for all of them.

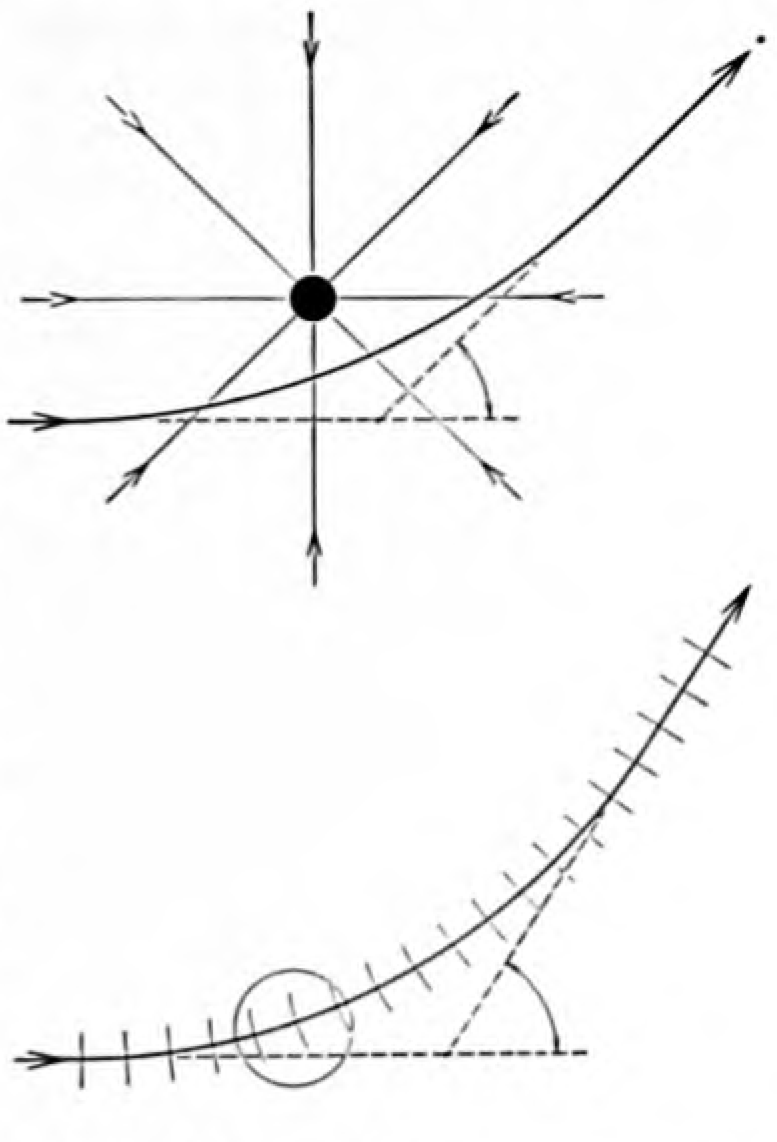
Probably the oldest model—and for some purposes still a very useful one—is the "spherical drop" or "liquid drop." Here the protons and neutrons (collectively called nucleons) which make up the nucleus are considered to be packed together like the molecules in a drop of water. On this model the nucleus has a uniform density everywhere in its interior, and a sharply defined surface. Furthermore, all nuclei, large or small, have the same density. Just as large drops of water contain more molecules than small drops but have the same density, so large nuclei have more nucleons, but these are no more nor less tightly packed than they are in small nuclei.

If this view is correct, there must be a rather simple rule governing the relative sizes of various nuclei. Their volumes obviously must be proportional to the number of nucleons they contain. And since the volume of a sphere depends on the cube of its radius, the radii of different nuclei must vary as the cube root of their numbers of nucleons. For example, if a large nucleus contains eight times as many protons and neutrons as a small one, it will have twice the radius.

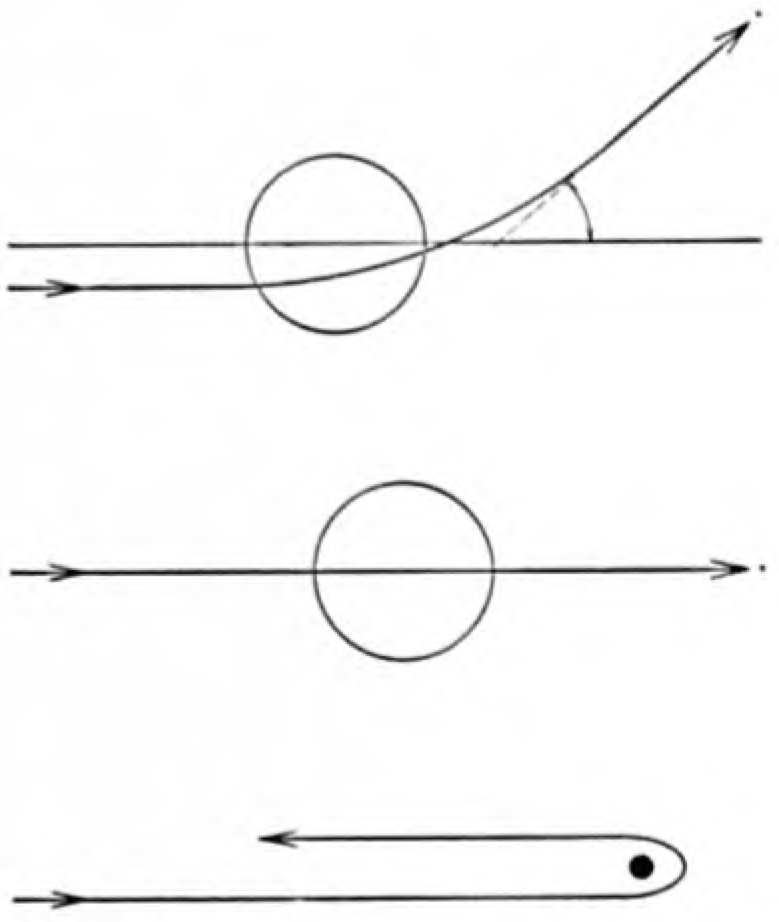
It is possible to go further and get a figure for actual as well as relative sizes. On the assumption of the spherical-drop model, various experiments indicate that the radius of a nucleus, measured in "fermis" (units of 10^{-13} centimeters), is 1.45 times the cube root of its number of nucleons. Thus the radius of the gold nucleus, which contains 197 nucleons, is $1.45 \times \sqrt[3]{197 \times 10^{-13}}$, or 8.45 fermis.

In addition to size and mass the nucleus has electric charge. This charge is positive, and is due only to the protons which the nucleus contains. The uncharged neutrons contribute to size and weight but not to electric charge. Now on the spherical-drop model the charge is also thought to be uniformly distributed throughout each nucleus. But the "charge density," that is, the amount of charge concentrated in a given volume, must vary from one nucleus to another, depending on the ratio of protons to total nucleons. The nucleus of ordinary hydrogen, which consists of just one proton and no neutrons, obviously has the highest possible charge density. In a nucleus where half the nucleons are protons (which is approximately the case for most light nuclei) the charge density will be half as great. In heavier nuclei the ratio of protons to total particles goes down to .39, so that the charge density is somewhat smaller.

There is no doubt that a number of important nuclear properties are reflected by the liquid-drop model. But there is also no doubt that an actual nucleus cannot be exactly like a liquid drop. It is extremely unlikely that the nuclear surface can really be sharp, with its density dropping from the constant interior value abruptly to zero. Modern quantum theory predicts that the density should



SCATTERING, or deflection, occurs whenever an electron passes through the force field of a nucleus. If it passes near the nucleus, its deflection results chiefly from the electric attraction between its negative charge and the positive nucleus (*top*). If it enters the nucleus, the situation is better pictured in terms of waves. The electron wave is refracted by the nuclear material (*bottom*), much as a train of light waves is refracted when it passes through a raindrop.



SCATTERING PATTERN depends on the structure of the target nucleus. An extended, diffuse nucleus (*top and center*) tends to give small deflections when the electron passes near its center. In fact, a particle passing exactly through the center is not deflected at all. A point nucleus, on the other hand, gives large deflections (up to 180 degrees). when an electron passes near the center (*bottom*). Thus relative degrees of scattering at small and large angles reflects some of the details of nuclear structure.

fall off to zero smoothly, from the high interior value through an outer layer or "skin." This has been realized for several years, but there seemed no way to find out how thick the skin was.

As a matter of fact, when one moves away from the simplified picture of the spherical drop, it is possible, in the present uncertain state of nuclear theory, to imagine a variety of models. Some calculations show that the nucleus may be a "soft sphere," whose density decreases steadily from the center outward. According to other theories the mass and charge may be concentrated in concentric shells. Some of these possibilities are illustrated in the drawings on these two pages. No one knew how seriously they should be taken.

Electron Probes

These questions were in the air in 1951 when the author began to think about a new way of examining nuclei. The idea was to shoot very high-speed electrons at them and see how the electrons were deflected, or, as the physicist says, scattered. Now scattering experiments are a classical technique of atomic physics. It was by observing the scattering of alpha particles that Lord Rutherford first discovered the existence of the nucleus. Later another British physicist, G. P. Thomson, used electron scattering to study the structure of molecules and atoms. In 1951 E. M. Lyman and his collaborators at the University of Illinois tried electron scattering on nuclei. With the moderate energies at their disposal, they were not able to make out any detail, but they did get an indication that heavy nuclei are somewhat smaller than had been thought.

To understand why high energy is necessary to reveal nuclear detail it is easier if we think of electrons as waves rather than as particles. Like all other subatomic bits of matter, electrons have wavelike as well as particle-like properties. (The rules of quantum physics tell us that the length of the waves depends on the energy of the particles; the higher the energy, the shorter the wavelength.) In many cases the behavior of electrons can be as well described from one point of view as from the other. For example, we can as well say that electron waves are diffracted by nuclei as that electron particles are scattered.

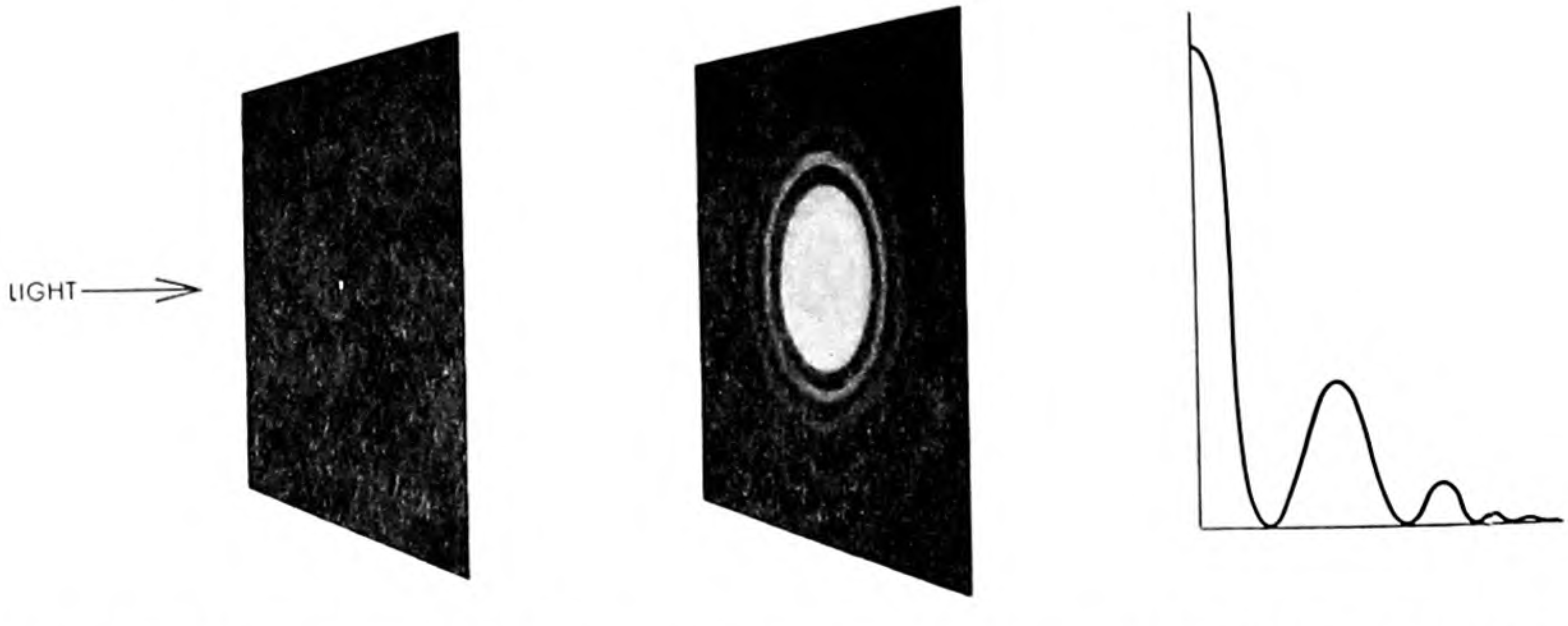
The electrons used by Thomson in his work on atoms had energies of a few tens of thousands of electron volts, which means that their wavelengths were on the order of 10^{-8} centimeters, which is 100,000 fermis. These waves cannot "see" the nucleus at all. Since they are about the same size as the atom's entire electron cloud, the nucleus in the cloud's interior will be entirely shielded from them. In Lyman's electron beam, at 15 million electron volts, the waves were hundreds of times shorter and could penetrate the cloud. But they were still considerably longer than the diameter of the nucleus, and hence could not show it in any detail.

But at Stanford in 1951 a great linear accelerator was being built that would produce an intense beam of electrons at energies approaching a billion electron volts. The corresponding wavelength would be measured in a few fermis. This is short enough to reveal nuclear structure in considerable detail. Thus it appeared that electrons could soon be used to examine the innermost part of the atom.

This was an exciting prospect. Until that time the chief nuclear probes had been protons, neutrons and alpha particles (which are made up of two protons and two neutrons). That is to say, the particles used to examine the nucleus were the same as those which compose it. This raised a difficult problem. One of the deepest mysteries that confronts physics today is the nature of the force that acts between nucleons. In a sense all of nuclear research is directed toward clearing up this fundamental question. Thus if we shoot protons or neutrons into a nucleus, we can interpret what happens only in terms of their interaction with the other nucleons. Yet this is the very problem we are trying to solve.

Electrons, on the other hand, are not nucleons; they are not subject to the mysterious nuclear force. When they pass near protons or neutrons, they are affected only by electric and magnetic forces, and these are as well known as any in physics. Calculations of electromagnetic interaction can be made with great confidence.

Let us consider in some detail what happens to electrons that are fired at nuclei, and what can be learned from their behavior. The electron may be



DIFFRACTION PATTERN obtained when light is passed through a small hole resembles the patterns of the electron-scattering experiments. The curve at the right shows how the intensity of light varies, starting at the center of the inner bright spot and moving outward across the pattern in a straight line. This is analogous to measuring the numbers of electrons scattered at various angles from the target. The dips in the scattering curves, although shallower than those in the light pattern, convey the same sort of information.

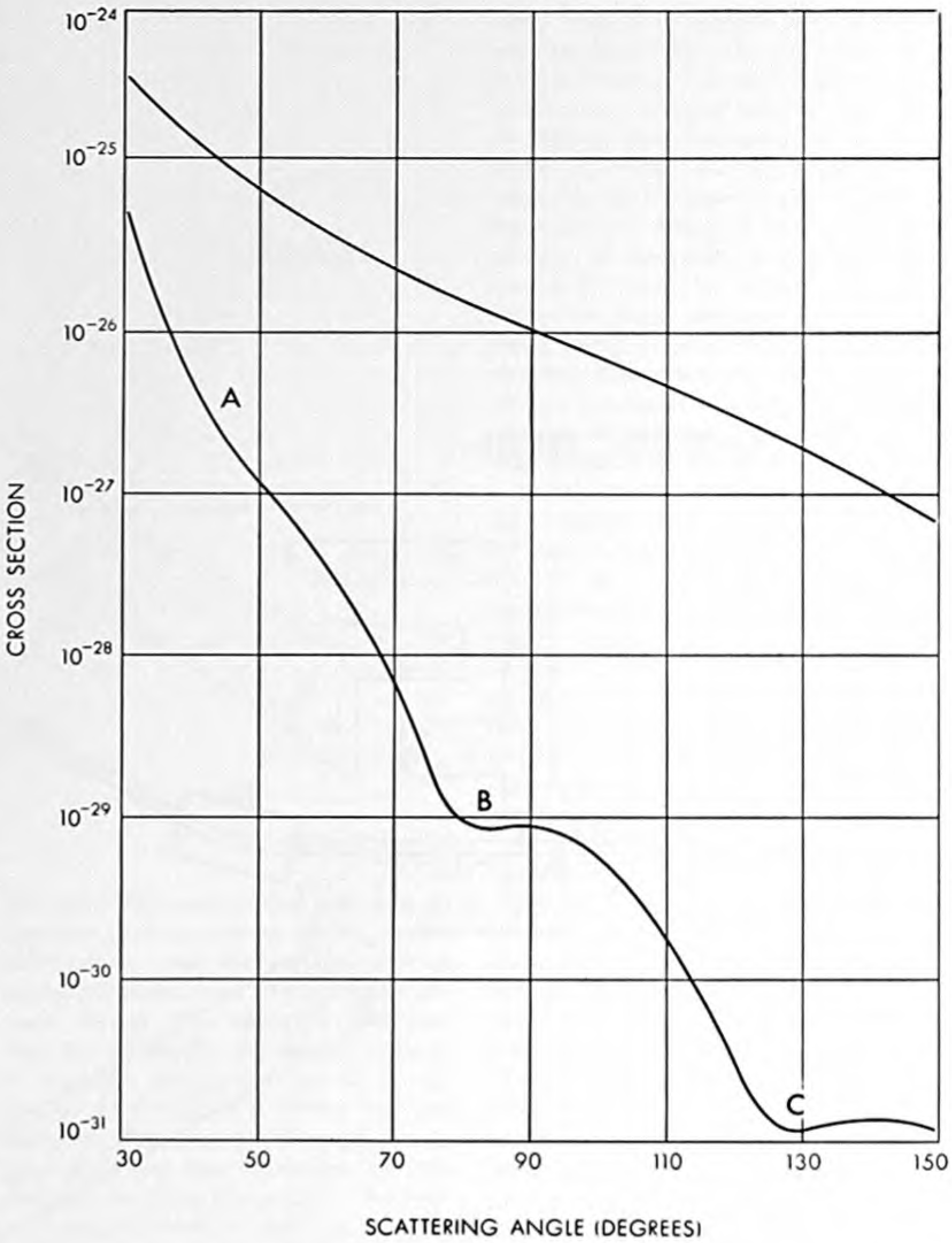
thought of as a negatively charged speck which, like the earth, is spinning on an axis through its center. The motion of the charge gives rise to a magnetic effect. Hence the electron is at the same time a tiny charge and a tiny magnet. A nucleus is a positively charged ball. It may or may not be magnetized depending on how the protons revolve within it. In some cases the motions give a net magnetic effect, in others they cancel each other out. For elements above an atomic weight of 10 the nuclear charge is so much greater than the magnetic strength, if any, that the latter is negligible. In this case the interaction between electrons and the nucleus is purely electrostatic. Lighter elements with magnetized nuclei exert both electric and magnetic forces on a bombarding electron, and to separate their effects is rather difficult. Marshall Rosenbluth, now at the Livermore Laboratory of the Atomic Energy Commission, has made the separation possible by calculating the pure magnetic scattering pattern.

In any case, an electron that passes through the force field of a nucleus is deflected. If it merely passes nearby, the situation can be pictured in terms of two attracting particles, as in the deflection of a comet by the sun. If the electron actually enters the nucleus, it is more convenient to think in terms of waves and diffraction rather than deflection [*see diagrams at top of page 135*].

The scattering pattern will depend on the nature of the target nucleus. If it is a point, or a small, densely packed sphere, then the closer a bombarding electron passes to its center the larger its angle of deflection. An electron passing very close to the target could be so strongly attracted that it would loop around and return in the direction from which it came. That is, its scattering angle would be 180 degrees [*see diagram at bottom of page 135*].

A diffuse or smeared-out nucleus would give a different result. An electron directed at the center of such a structure would see as much positive charge on one side of its path as on the other. Hence it would not "know which way to turn," and would pass straight through.

Here we have come to the heart of the electron-scattering method. With a dense, tightly packed nucleus we expect a considerable amount of scattering at large angles, up to 180 degrees. With a diffuse, "soft" nucleus the large-angle or backward deflection will be very much reduced in favor of forward scattering. The two curves in the diagram above show what is expected theoretically

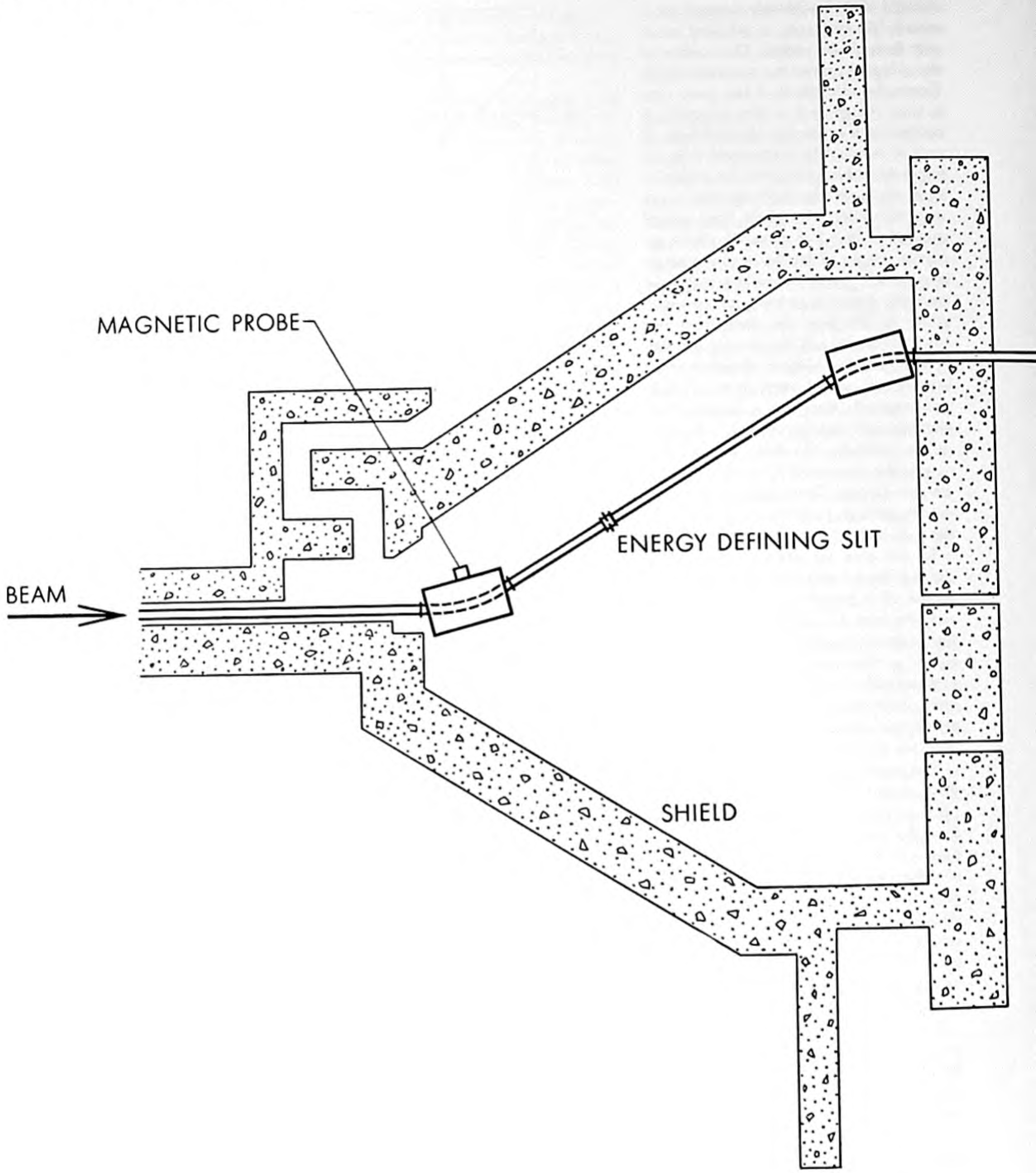


THEORETICAL CURVES show the scattering patterns expected from a point charge (*top*) and a uniform soft cloud (*bottom*). The vertical scale is a measure of the percentage of the incoming electrons which would be detected at the various angles on the horizontal scale.

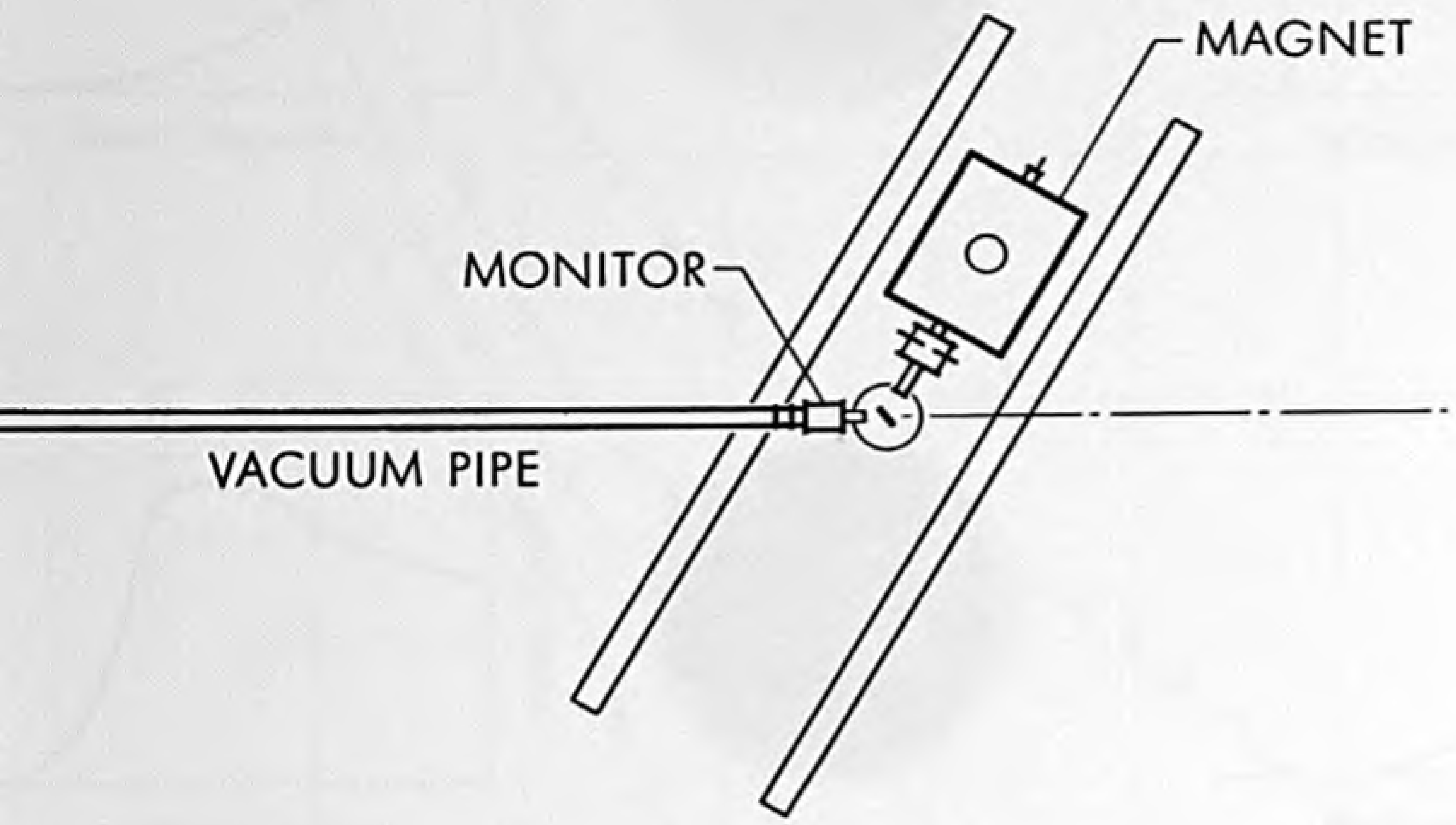
when electrons are scattered from a point charge and from a uniform soft cloud. It is also possible, although in some cases very difficult, to compute the scattering pattern for more complicated models, as we shall soon see.

A reduction in the deflections at large angles is not the only effect of a soft nucleus. Its scattering pattern is also marked by a series of ups and downs or diffraction "wiggles." To appreciate their significance we may compare the diffraction of electrons by the nucleus with the diffraction of a beam of light when it passes through a small circular hole in an opaque barrier. A screen on which the light is allowed to fall after passing through the barrier shows a bright spot directly opposite the hole, surrounded by alternating dark and light rings, the light rings growing fainter as they get bigger [*see diagram at the bottom of page 136*]. If this pattern were translated into a graph showing brightness at various angles from the forward direction, the result would be a curve with diffraction wiggles resembling those in the electron-scattering curve, although more pronounced.

From the spacing of the rings or wiggles in the light-diffraction pattern it is possible by the methods of theoretical optics to figure out the diameter of the hole. Similarly, the spacing in the elec-



EXPERIMENTAL ARRANGEMENT for electron-scattering measurements is shown diagrammatically. The magnetic probe and energy-defining slit pick out a narrow energy band from the incoming electron beam. Electrons striking the target are picked up and focused onto the detector (*not shown*) by the magnet, which can be moved to different angular positions. The monitor records the total number of electrons that pass through the target material. A 10-ton shield around the detector cuts out background radiation.



tron patterns gives information about the size of the diffracting nucleus.

The analogy is not quite exact. To make it better we should have to imagine that the hole in the light experiment contains a lens made of nonuniform glass, so that the refractive index is not the same throughout the lens. The effect of such a lens would be to distort the normal pattern. The analogous problem would then be to figure out from the irregular diffraction rings both the size of the hole and the exact make-up of the lens. In the actual case we must calculate the size of the nucleus and its internal distribution of charge. As may be imagined, this can be extremely difficult. Without the aid of high-speed computers it would in many cases be impossible. However, it has now been done for a large number of nuclei, and a striking new picture has emerged.

The Apparatus

Before turning to the results, let us see how the diffraction experiments are made. The Stanford linear accelerator shoots electrons at a rate of 60 bursts per second; each burst lasts a millionth of a second and contains about 10 billion particles. The machine will accelerate electrons to an energy of 700 million electron volts, but our analyzing apparatus cannot handle particle energies over 550 Mev. For reasons that will soon be apparent, the conditions of the experiment require that the particles striking the target nuclei have a very sharply defined energy. The electrons emerging from the accelerator are already at nearly the same energy—the spread is only five to 10 Mev. This is further reduced by letting them pass first through a magnetic field and then a narrow slit [*see diagrams on these two pages*]. The magnet bends electrons of different energies in different directions and the slit picks out those that are traveling in one direction.

The narrow beam of uniform-energy electrons is then directed against the target material—for example, a gold foil about two thousandths of an inch thick. To determine the scattering patterns, one might suppose that it would only be necessary to move a detector around the foil and count the number of electrons deflected at various angles. It would be nice if this were so. It would make our apparatus some $200,000 cheaper and 45 tons lighter. The weight and the money represent an enormous D-shaped magnet which sorts out the scattered electrons by energies.

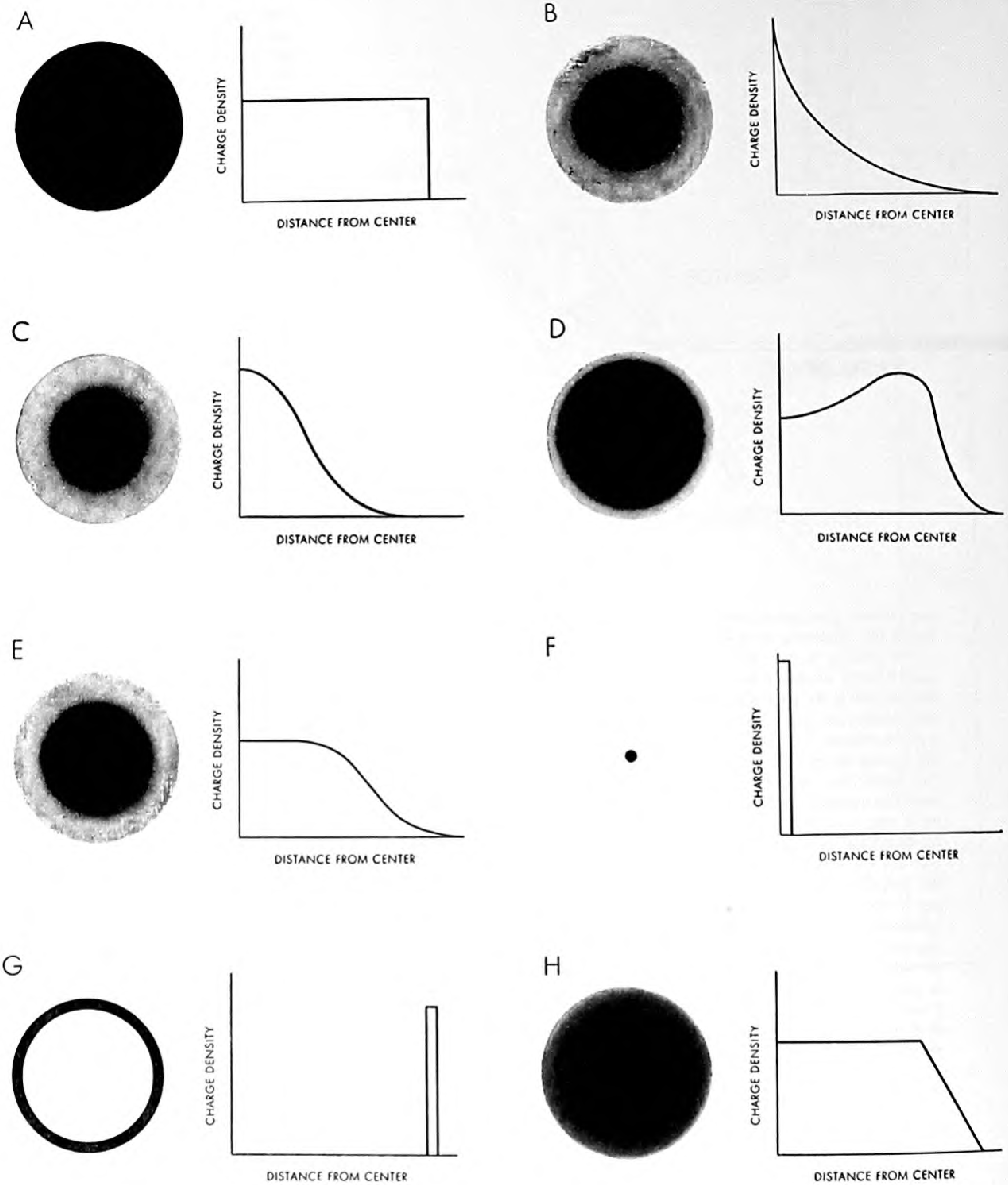
Why is this necessary? Why not, in fact, simply count the total electrons scattered at each angle? The reason is that all the particles deflected at a particular angle have not undergone the same type of interaction with a target nucleus; hence they do not convey the same information. In some collisions the electron and nucleus behave like a pair of billiard balls bouncing off one another, or, rather, like a ping-pong ball bouncing off a cannon ball. That is to say, the total energy of motion (kinetic energy) of the particles after the collision is the same as before. These are known as elastic collisions. A large nucleus, being so much heavier than the electron, does not recoil appreciably, so that in an elastic collision the electron bounces off with just about the same energy it had on its approach.

In other cases, the electron gives up some energy which is not accounted for by recoil motion of the struck nucleus. That is, the energy exchanged does not remain kinetic. Instead the nucleus becomes "excited" from its normal or "ground" state to a state of higher internal energy. (We may crudely picture what happens by saying that the individual nucleons go into more energetic motion.) When this happens, the collision is said to be inelastic.

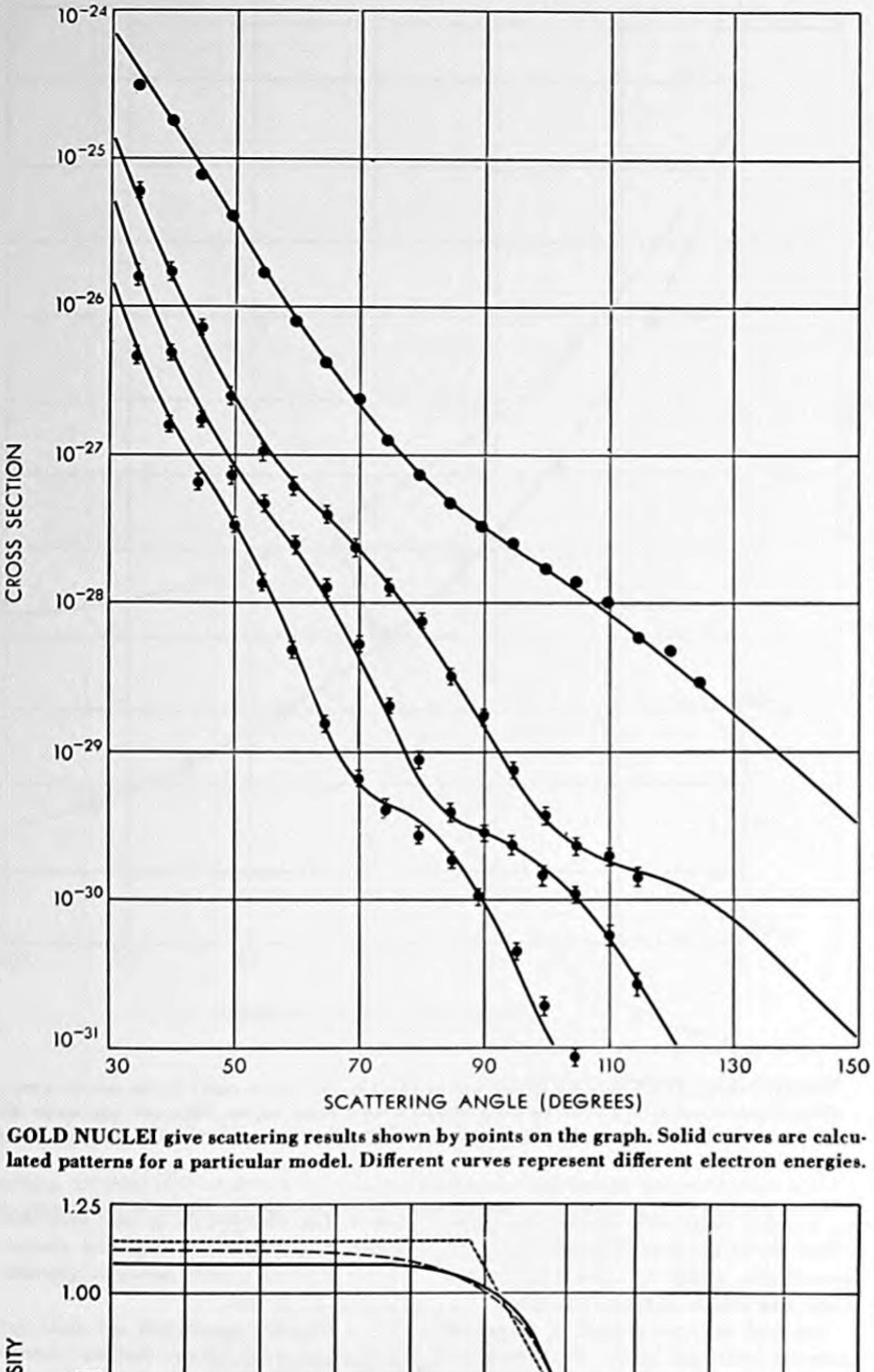
If we are interested in examining the nucleus in its normal condition then the electrons scattered inelastically are no help. They have "seen" the nucleus in an excited state. Hence we wish to pick out the elastically scattered electrons—the electrons emerging from the collision with the same energy they had going in. This is what the magnet does. Scattered electrons enter the semicircle on one side of the center and are bent 180 degrees by the magnetic field so that they emerge at the other side. Particles of different energies follow different paths, so that it is possible to focus those of a particular energy at the detector. The reason the magnet must be so big and powerful is that high-energy electrons are hard to deflect. The first apparatus we built had a magnet weighing two and a half tons and could handle electrons only up to 190 Mev. The new device, in use for about a year, can force 550 Mev electrons around its semicircular track. It is also sensitive enough to select an energy band only .8 Mev wide at 400 Mev. In other words, it can separate electrons whose energies differ by one part in 500.

Our detector is a small piece of lucite that glows momentarily when a fast electron passes through it. A photomultiplier tube picks up the light flashes and produces a corresponding series of electric pulses which are fed into a counter.

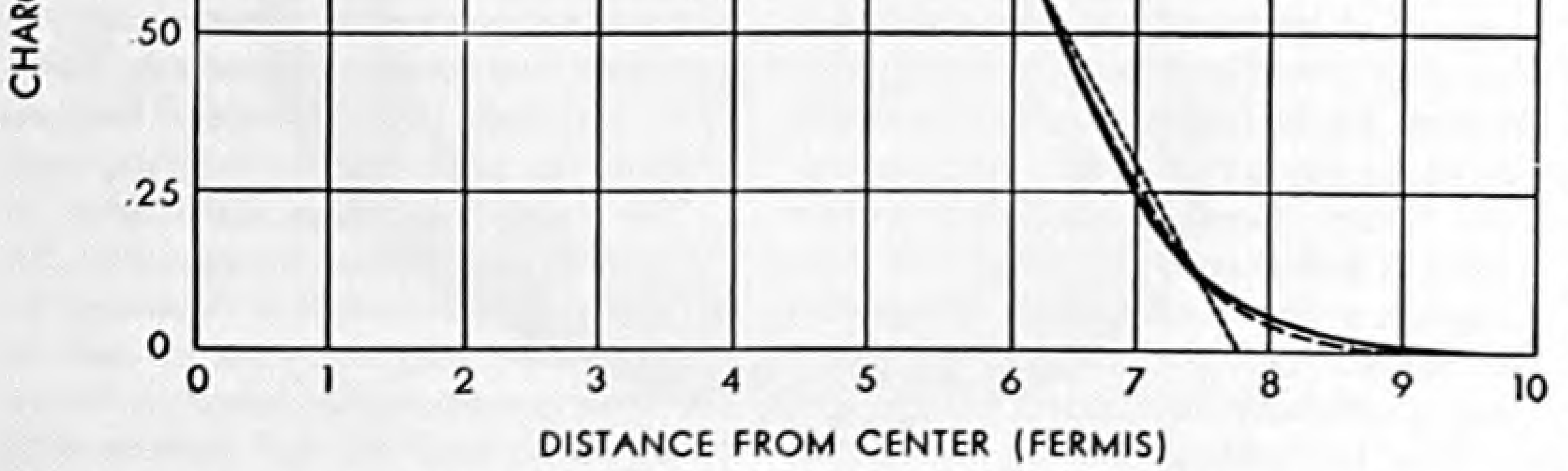
The actual number of scattered electrons counted at any angle is, of course,



MODELS OF THE NUCLEUS, showing various conceivable distributions of electric charge within it, are illustrated here. The graphs show how the density of charge varies from the center of the nucleus outward for each model. To the left of each graph is an imaginary rendering of what the model would "look like" in cross section. Model A is the liquid drop, with its constant density and sharp boundary. Other possible structures with sharply bounded surfaces are the point nucleus (F) and the shell (G). The electron-scattering experiments of the author and his colleagues indicate, however, that the nuclear boundary is not sharp. Their results suggest a skin, the density of which falls off gradually. Models such as E or H give the best agreement with the experimental findings.



GOLD NUCLEI give scattering results shown by points on the graph. Solid curves are calculated patterns for a particular model. Different curves represent different electron energies.



MODEL OF GOLD NUCLEUS which gives close agreement with experiments is shown as a solid curve. Dotted lines represent alternative models that might give nearly the same results.

of no significance unless the total number of incoming particles is known. What matters is the fraction of the total that is scattered in a given direction. Each incoming electron knocks a burst of secondary electrons out of the plates of a monitor material, and a count of these bursts indicates the number of particles in the incident beam.

The magnet, the detector and a 10-ton lead and concrete shield which surrounds it are all mounted on an obsolete five-inch naval gun base provided by the U. S. Navy. On this movable platform the apparatus can be swung to various angles around the target. A remote-control arrangement can position the gun base to an accuracy better than a tenth of a degree.

Our apparatus operates in many ways like an optical system, and has even been called a nuclear microscope. Just as a lens collects light scattered by an object, the magnet collects electrons bouncing off a target. The lens focuses the collected light to a spot; so does the magnet in the electron analogy. However, the magnet does more, since it sorts electrons into separate energy ranges. The optical analogy would be a spectroscope or spectrometer, sorting colors or wavelengths of light. The magnet performs the operations of collecting, bending and refocusing the electrons just as a spectroscope collects light waves, separates their various colors with a prism or diffraction grating, and refocuses them.

New Nuclear Models

Let us now look at some of the results that have been obtained with this apparatus. The experimental procedure is, as has been pointed out, quite simple in principle. We simply set the accelerator for a given energy, and count the percentage of elastically scattered electrons at various angles around the target. The dots in the chart at the top of the next page show the result of a number of such runs on gold nuclei at energies ranging from 84 to 183 Mev.

While the experimental group was making these measurements, a team of theoretical physicists including D. R. Yennie, G. D. Ravenhall and R. N. Wilson was busy calculating the expected diffraction pattern for various nuclear models such as those illustrated on page 140. One specific model of the gold nucleus has a dense core extending about four fermis from the center, and then a rapidly thinning "skin" which drops away to nothing at around nine fermis [*see chart at bottom of this page*]. The theoretical diffraction pattern at various energies from 84 to 183

Mev for this model are also shown on the chart at the top of page 141 as solid curves. The agreement between the experimental and theoretical curves is nothing short of astonishing. Evidently the distribution of charge in the gold nucleus must be very much like the one in this model. Since protons and neutrons are presumably distributed in the same way, this distribution should also apply to the total mass of the nucleus.

If we define the skin thickness of the gold nucleus as the distance between the point where the charge density is 90 per cent of the maximum and the point where it has fallen to 10 per cent, we find that the thickness is close to 2.4 fermis. Taking as a measure of nuclear size the distance from the center to the point where the density is 50 per cent of the maximum, this turns out to be approximately 6.3 fermis.

Now, when we turn to investigate other nuclei, a surprising regularity appears. Together with Ravenhall and Beat Hahn, the author has made a systematic investigation of selected nuclei from mass number 40 to 238. Throughout this entire range the skin thickness is 2.4 fermis! The size of the dense inner core varies, but the fuzzy outer layer is the same thickness for all these nuclei.

As a measure of nuclear size it is more convenient for many purposes to use another average value for the radius rather than the distance from the center to the 50 per cent density point. This average is known as the "root mean square." When it is used, the value for the radius of a heavy nucleus obeys a simple law: its value in fermis equals 1.18 times the cube root of the mass number. This rule is reminiscent of the cube-root law for the liquid-drop model, but it implies a smaller nucleus than that suggested by the older view.

In nuclei below mass number 40 we find that the inner core practically disappears, and that the density decreases from the center out. These lighter nuclei obey a slightly different size rule: the root mean square radius is 1.35 times the cube root of the mass number.

At just about the time when we first obtained these results Val Fitch and L. James Rainwater at Columbia University were measuring nuclear sizes (but not skin thickness) by an entirely different method. They found the same law as we did for nuclei above mass number 40. For the lighter nuclei, however, their results were 1.18, as against our 1.35.

The Proton

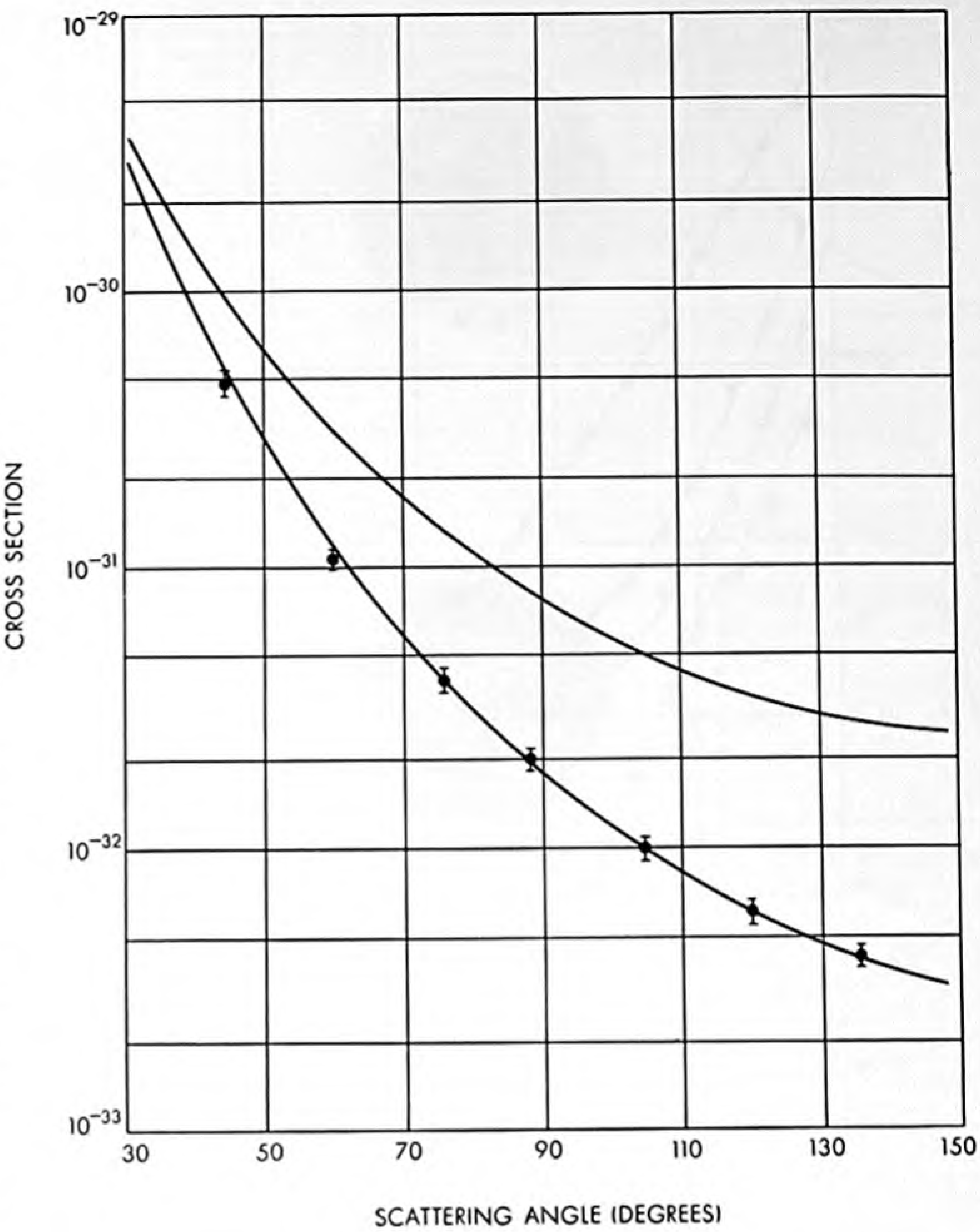
Having proved that our electron beam could indeed see into nuclei, we began to wonder about still smaller particles. What about the proton itself? Is it a dimensionless point? Or does it too have a finite size and an internal structure?

To find out, we placed a target of gaseous hydrogen in the electron beam and again proceeded to measure elastic scattering. As the chart on this page shows, the results were quite clear. The amount of backward scattering is much less than would be obtained from a point proton. Again, one particular theoretical curve fits the actual results very closely. The charge distribution which gives this curve is bell-shaped or "Gaussian" [*see diagram at top of next page*]. It can also be proved that the proton's magnetic field is similarly distributed.

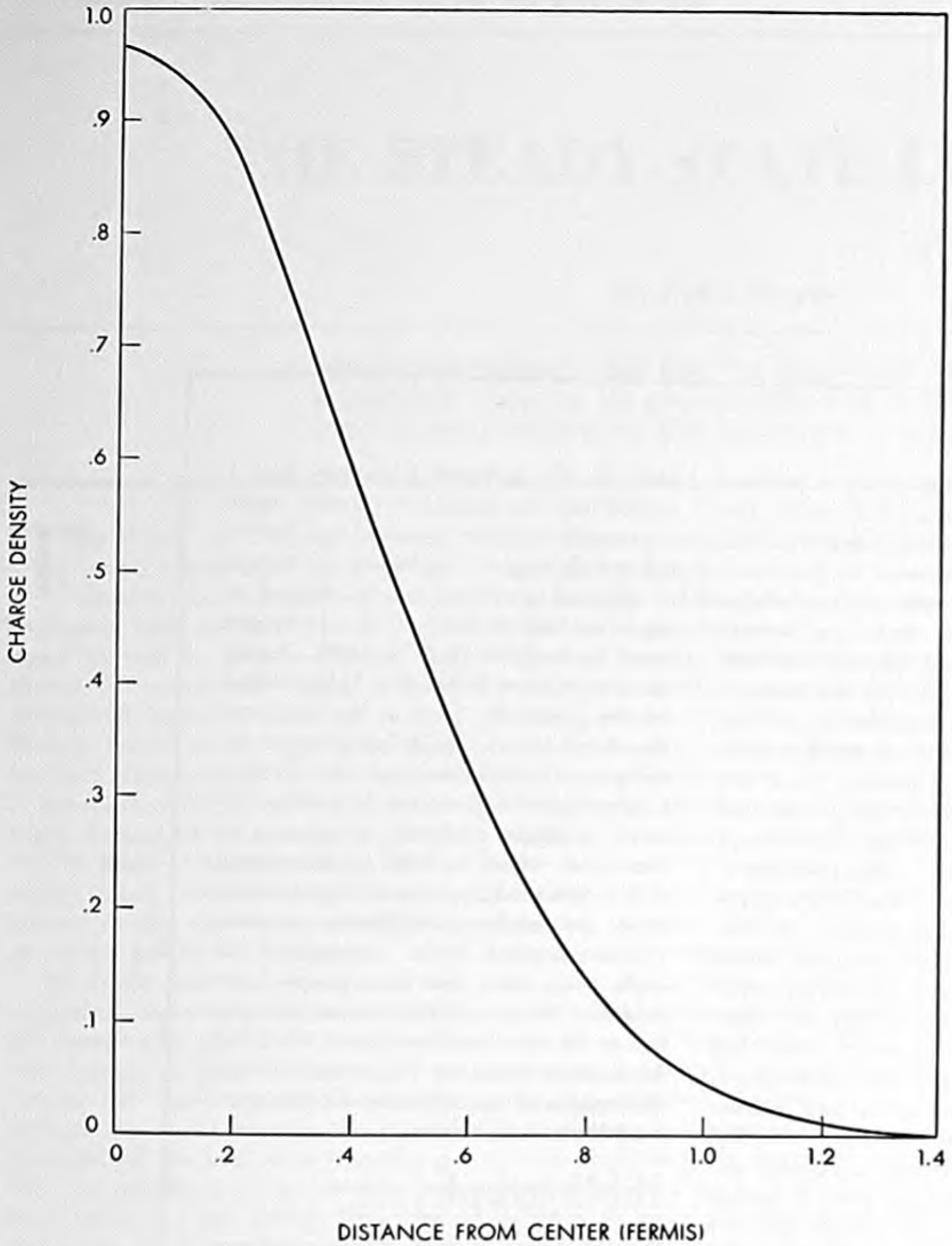
Thus the proton must be considered an extended body, and our electrons have for the first time actually seen inside it. The charge falls to zero only at a distance of 1.4 fermis from the center. The root mean square radius is approximately .75 fermis.

The proton experiments are quite recent, and we cannot say that the charge distribution shown is absolutely the only one that would give a close agreement with the experimental results thus far. But we feel that further experiments will specify a true model not basically different. It is likely that the central features will change more than the outlying ones.

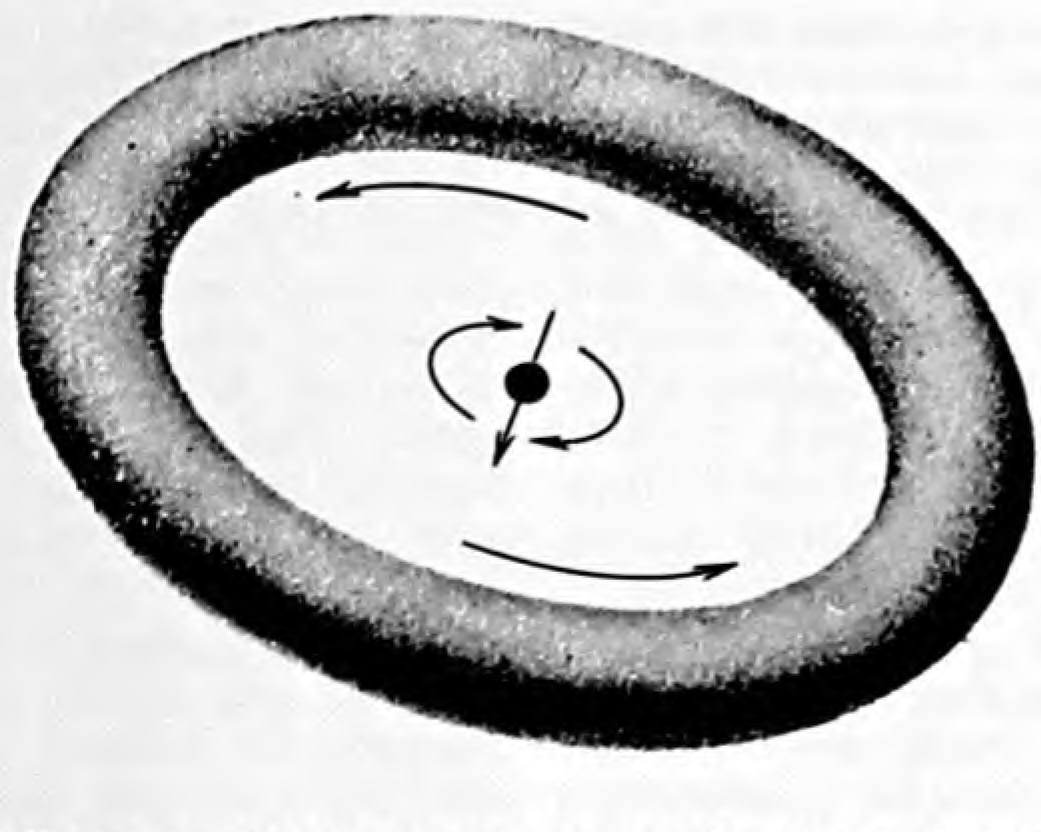
The extended charge distribution of the proton may explain the apparent discrepancy between radii as measured by electron-scattering experiments and by the older measurements based on the interaction of neutrons and protons with nuclei. If a gold nucleus and a proton are placed with their centers separated



PROTON-SCATTERING PATTERN would be as in the upper curve if the proton were a dimensionless point. If its charge were spread over a finite region, then one particular distribution would give pattern of lower curve. Points show actual scattering measurements.



MODEL OF PROTON shown above gives a theoretical scattering pattern very close to the one observed. The density of charge falls off from the center outward in a "Gaussian" curve.



MESON THEORY suggests that the proton may actually consist of a spinning "bare nucleon" which is essentially a point, surrounded part of the time by a rotating meson cloud.

by 8.45 fermis (the old value for the gold radius) there will be a considerable overlap of charges. The outer portions of the skins are already in contact. Thus the proton finds itself, at this distance, in a situation not radically different from that of an outer proton in the gold nucleus. There is no apparent reason why it should not interact strongly with the other nucleons of gold. Hence we may expect that a radius measured by nuclear methods that involve the strong nucleon interaction will appear to give a larger value than the electromagnetic radius measured with electrons.

According to present theory, the model of the proton obtained from the scattering experiments may not really represent a single, smeared-out particle. Instead, the proton may actually consist of a pointlike "bare nucleon" intermittently surrounded by a cloud of mesons [*drawing at lower left*]. It is probably the meson cloud that we are probing.

The theory says that the proton erupts from time to time, emitting a meson which whirls about for an unimaginably short period and then is sucked back into the proton again. The process of emission and reabsorption is considered to be an ever-present, essential activity of the proton (and the neutron as well). One problem has been to decide what fraction of the total time the meson spends outside the proton. Our measurements can be interpreted as indicating that the fraction is a few tenths or more. This is a higher value than had been previously estimated.

It is thought that the mysterious nuclear force arises from an exchange of mesons between nucleons. If electrons can be used to "see" the mesons, they may help clear up the mystery.

As this is written our group is busy with new scattering experiments. We are refining our observations on the proton. Preliminary investigations of the alpha particle show that it has a charge distribution like the proton's and is unexpectedly compact. It is only a little larger than its two protons together, despite the fact that it also contains two neutrons. Experiments with the deuteron (the heavy hydrogen nucleus, containing one proton and one neutron) show that it is bigger than the alpha particle. The deuteron observations may also give some information about the distribution of the neutron's magnetic field. It may soon be possible to tell whether the neutron and the proton are, as current theory says, alike except for their charge.

This is only a partial list of the exciting problems that are waiting to be investigated with high-energy electrons.

The Author

ROBERT HOFSTADTER is professor of physics at Stanford University. He grew up in New York City, and while in high school was interested in literature and philosophy. "On entering the College of the City of New York," he writes, "I found that although physics was less alive than literature, the physics instructor was much more stimulating. At his suggestion I took some advanced mathematics and physics. I liked to be at the source of things, and felt that physics was fundamental to everything else, except possibly mathematics. I was also stimulated by the lives of some of the great physicists and mathematicians. At this time I felt that the laws of physics could be tested and those of philosophy could not. Halfway through college my mind was made up that I wanted to teach and do research in physics." Hofstadter graduated from City College in 1935 *summa cum laude*. A Coffin fellowship, awarded by the General Electric Company, enabled him to do graduate work at Princeton University. "I was pushed into experimental work," he recalls, "by the Coffin requirement that a man must do research even in his first year. In my second year there seemed to be an open place in the infrared laboratory and I moved into that branch of physics, again not exactly because I wanted to." Studying simple organic molecules by means of infrared spectroscopy, he helped to elucidate the nature of the hydrogen bond; he took his Ph.D. in 1938. During the first years of World War II he worked on the proximity fuze at the National Bureau of Standards; he later helped develop servomechanisms at the Norden Laboratories Corporation. In 1946 he became assistant professor of physics at Princeton, where in 1948 he discovered that sodium iodide activated by thallium made an excellent scintillation counter. "I was extremely lucky. Throughout the eight years since that time people have searched intensively for a better material but so far none has been found." In 1950 he went to Stanford, where he built the first magnetic spectrometer for the Stanford linear accelerator.

Bibliography

THE DISTRIBUTION OF CHARGE IN THE NUCLEUS. Kenneth W. Ford and David L. Hill in *Annual Reviews of Nuclear Science*, Vol. 5; 1955.

SCIENTIFIC AMERICAN September 1956, Vol. 195, No. 3, pp. 157-166 OFFPRINT 218

THE STEADY-STATE UNIVERSE

by Fred Hoyle

Some cosmologists hold that the large-scale features of the expanding universe do not change, and that its density is maintained constant by the continued creation of matter.

The theory of a steady-state universe leads to many startling conclusions: that the universe had no beginning and will have no end, that space as well as time is infinite, that matter is continually being created throughout space—to mention a few. Human nature being what it is, there has been a tendency to become involved in emotional attitudes toward these concepts, instead of confining the discussion to purely scientific criteria. If the writer, along with critics, has transgressed in this respect, he promises to give some redress in this article.

The steady-state theory holds that the large-scale features of the universe do not change with time. Only the galaxies and clusters of galaxies change; if we "smear out" their material uniformly through space and consider the general properties of the cosmos, it is unchanging. The expansion of the universe is a basic feature of the theory. The question arises: If the galaxies are moving apart from each other, why does space not become more and more empty? The answer of the theory is that new galaxies and clusters of galaxies are constantly being formed, their rate of formation just compensating for the separating effect of the expansion. So a stable situation is preserved.

Before we go on to consider the reasoning, predictions and tests of steady-state cosmology, the writer should point out that his own approach to the theory, and also that of William H. McCrea of the University of London, differs rather markedly from the approach of Hermann Bondi and Thomas Gold. The writer's approach is a mathematical one developed in the framework of the theory of relativity. Bondi's and Gold's is founded on an intuitive but powerful physical principle. To understand their outlook we must look into the nature of this postulate, which is called the "cosmological principle."

Cosmology differs from all other branches of physical science in a very important respect. Whereas other physical scientists deal always with isolated systems, whose "boundary conditions" can be defined, a cosmologist has to deal with a nonisolated system. To cope with this unhappy situation he is forced to adopt a "symmetry" postulate, which says that, local fluctuations apart, the universe will look the same from wherever one views it. That is to say, it assumes that observers attached to different galaxies anywhere in the cosmos would all obtain exactly the same large-scale picture of the universe. But if the universe changes with time, this implies that the different observers compare their pictures at the same time, which of course requires us to have a definition of what we mean by "at the same time." In order to make a definition of simultaneity possible, the mathematician Hermann Weyl advanced the additional postulate that the motions of the galaxies follow a regular type of pattern, whose exact nature need not be described here.

Instead of this additional postulate Bondi and Gold proposed a single all-embracing "cosmological principle": namely, that the large-scale features of the universe are the same not only from every point of view in space but also from every point of view in time. This symmetry hypothesis leads immediately to the conclusion that the universe is in a steady state. It is then immaterial whether the observers compare their pictures "at the same time" or not.

The outlook of Bondi and Gold has a compelling simplicity. Moreover, symmetry postulates have repeatedly demonstrated their power in theories of physics during the present century (*e.g.*, the positive and negative particles of nuclear physics). But to my own taste it is preferable to start with a mathematical definition of the continuous creation of matter within the framework of the relativity theory and then to derive the steady-state solution as a consequence of field equations.

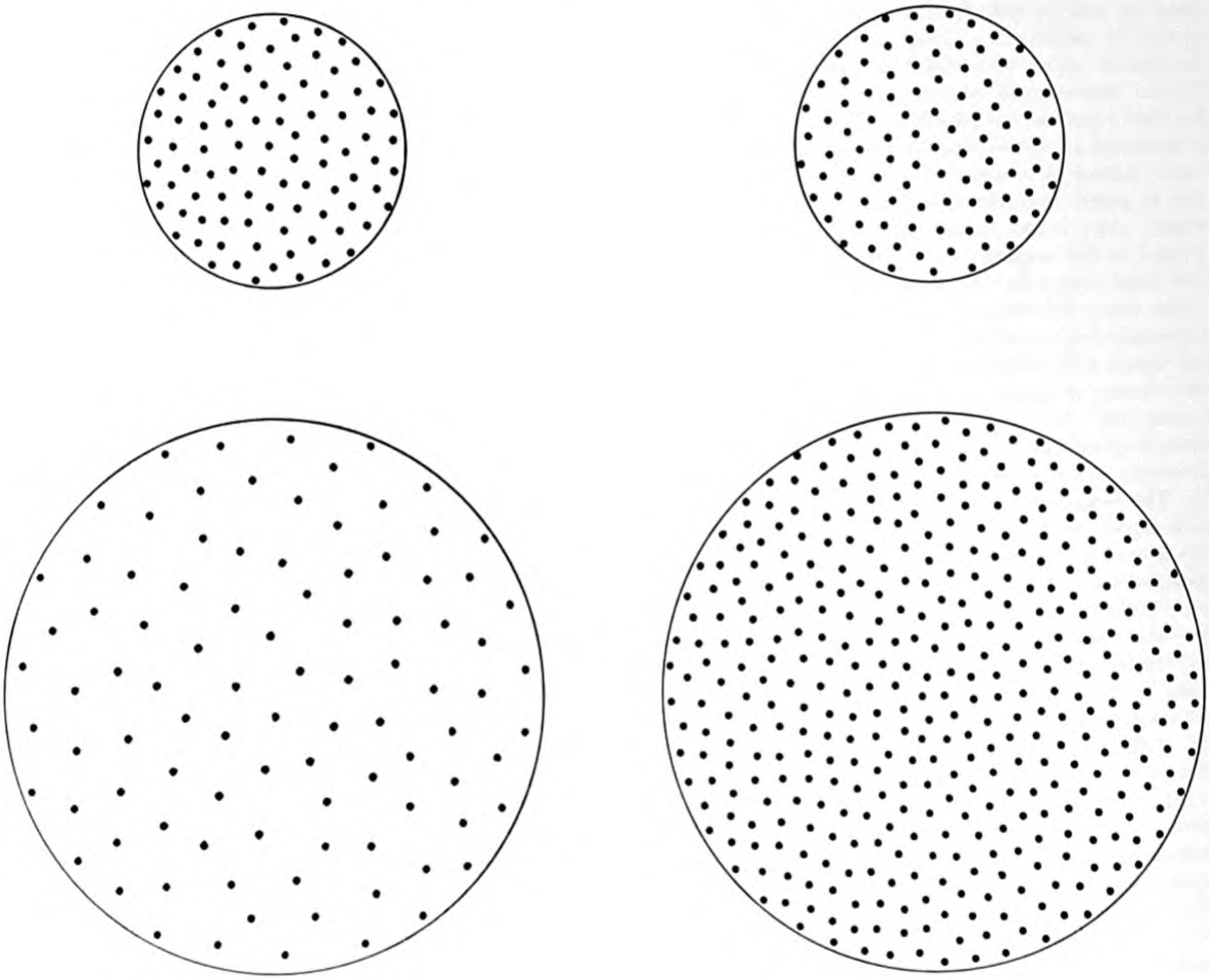
At first sight the creation of matter may seem a queer concept to be invading scientific thought. But as other articles in this issue make abundantly clear, the origin of matter must enter all cosmologies. Nowadays we are coming more and more to realize that hydrogen is the original material—the material out of which the other elements have been produced by nuclear reactions inside stars. This transmutation of hydrogen is going on all the time.

Why is there any hydrogen remaining in the universe? Why was it not all used up in the production of heavy elements eons ago? If we assumed that the hydrogen of the universe has existed for an infinite time, there would be two conceivable answers. We might suppose that the hydrogen has not had sufficient time to become transmuted into other elements because the stars were born only recently, that is, within the last five billion years or so. But it would follow from this that the hydrogen remained stable for eons of time and then suddenly five billion years ago began to condense into stars and galaxies. This seems less than plausible. The other possibility, assuming the hydrogen is infinitely old, is that we still find it on hand because the higher elements formed from it break down to hydrogen again. The chief objection to this idea is the difficulty of explaining how the energy necessary for the breakdown would be supplied. Decomposition of the heavier elements into hydrogen requires absorption of energy—the reverse of the release of en-

ergy that occurs when hydrogen nuclei combine. To provide an amount of energy adequate to account for a sufficiently large-scale reconversion of the heavier elements, nothing less than an implosion of the whole universe (as opposed to an explosion) apparently would suffice.

We are thus led to the conclusion that the hydrogen we observe is nct infinitely old: it has originated within some finite time and has not yet been converted to heavier elements. Both the evolutionary and the steady-state theories of the universe agree on this point But there the similarity between them ends. The evolutionary theory argues that all the hydrogen was created in an explosive beginning some five and a half billion years ago. The steady-state hypothesis holds that hydrogen has been created at a steady rate throughout infinite time and is still being created at the same rate today.

If hydrogen has been present for an infinity of time, and has steadily been converted to heavier elements in stars, why don't we see galaxies made of very old matter? Why do we see only comparatively young galaxies, composed almost entirely of hydrogen? The answer of the steady-state theory is that the expansion of the universe spreads galaxies apart as they age, and the old material is rapidly diluted, in terms of its mean density in the universe as a whole. Meanwhile new hydrogen, and new galaxies, are just as rapidly being created. According to the mathematics of the theory, the expansion of the universe and the creation of new material go on at rates such that the mean density of 200-billion-year-old material, for example, is less than that of recently formed material by a factor of 10^{43} (1 followed by 43 zeros). It must be emphasized that this figure is a mean averaged over the universe as a whole: it does not apply to individual galaxies or clusters of galaxies. Expansion takes place in space *between* galactic systems: the individual galaxies and clusters do not themselves expand. The very old material of the universe is concentrated in very old galaxies. By virtue of the universal expansion these are now extremely far apart. Possibly there are some moderately old galaxies within the range of our telescopes. If a method could be worked out to identify distant galaxies composed of comparatively old matter,



EVOLUTIONARY AND STEADY-STATE VIEWS are compared in these diagrams. At left is a schematic view of an evolutionary universe. At the top is a sample of the universe, with the galaxies represented by dots. At the bottom is a picture of the same galaxies after the passage of time. The galaxies have merely receded from one another. At right is the same kind of view of a steady-state universe. At the top is a sample of the universe. At the bottom is a picture of the volume occupied by the same galaxies after the passage of time. In that time, however, new galaxies have been created, maintaining the density of the galaxies as before.

it would provide a test of the steady-state theory.

Approaching the steady-state theory from the mathematical point of view, our first step evidently must be to construct a mathematical law representing the origin of matter. We wish to formulate this law within the logical framework of the theory of relativity: like the evolutionary theory, steady-state cosmology makes use of the powerful equations devised by Albert Einstein to describe the four-dimensional space-time continuum. We can indicate briefly here some of the main principles involved, though the equations themselves are too complex to examine in detail.

The theory of relativity begins by generalizing the ordinary laws of motion in three-dimensional space to describe the properties and the non-Euclidean geometry of the four-dimensional space-time field. These laws can be set down in four equations: one equation for the law of conservation of energy and three for the conservation of momentum. Our problem is to frame the law of origin of matter in such a way that it can be introduced into these four conservation equations.

As a first step we must define energy and momentum, for the theory of relativity does not itself define them. It is most reasonable to choose definitions which will yield equations as closely analogous as possible to the ordinary equations describing the laws of conservation in our familiar (Euclidean) world. The evolutionary cosmologists seem at first sight to have done this, but it turns out that their conservation equations do not contain any generalized analogue of certain terms, known as "fluid stresses," which play a part in the ordinary equations. Now when we define energy and momentum in a way that yields such a generalization, the outcome of the equations is a steady-state universe, not an evolutionary one.

The equations, so generalized, imply a "feedback" relation between the expansion of the universe and the origin of matter. If the expansion rises above a certain critical rate (related to the rate of origin of matter), the feedback slows the expansion. If the universe's expansion slows down to less than the critical rate, the feedback speeds it up. Thus the interaction between the expansion and the creation of matter maintains a steady state in which the mean density of matter in the universe remains constant.

To many people the notion of continuous creation of new matter in space seems an outright violation of the conservation of energy. But this indicates a confusion between a closed system and the very different situation in an open system. The theory of relativity says that in an open, infinitely expanding universe, local concentrations of energy are related to the energy of expansion of the whole universe. The energy of expansion can take a form leading to a continuous creation of local matter.

The same question that is asked about the creation of matter might be asked about the red-shift of light from distant galaxies. The reddened light is weaker than when it started on its journey. Where does the lost radiant energy go to? It goes into a slight increase in the rate of expansion of the whole universe. The point is that for a total reckoning of the conservation of energy and matter in the cosmos we must take the expansion of the universe into account. We cannot balance the energy books strictly and completely within the confines of any locality, because no locality in the universe is entirely closed.

Before we drop this issue it is perhaps worth noting that we can consider the conservation question in purely operational rather than theoretical terms and come out with the same result. Suppose observers on the earth measured the energy content of a given portion of the universe, say that within the reach of the 200-inch telescope, and suppose this was done on several occasions at widely separated times. If the conservation of energy is to hold, the measured energy content must remain unchanged from one occasion to another. This would be true in a steady-state universe, but not in an evolving one. Furthermore, in a steady-state universe conservation in this operational sense holds good for an observer in any galaxy.

The two features of the steady-state theory that seem to cause the greatest general surprise are (1) that the theory possesses a clear-cut mathematical basis, and (2) that the theory is highly susceptible to test by observation. How can it be tested? Obviously we cannot test it in the laboratory—unless we were to find some way to speed up the creation of matter artificially—for the rate of creation, according to the theory, is negligible in terrestrial terms: in the space of the average physics laboratory one new hydrogen atom would materialize in about 1,000 years. But on the cosmological scale there are many possible tests.

Firstly, at the farthest range of our telescopes we are seeing galactic systems as they were a billion or more years ago. Hence information can be obtained about how things used to be in the past, and this information can be compared with the cosmic scene close by us in space and time. Since the steady-state theory requires that there be no difference in large-scale properties between the past and the present, the theory is clearly exposed to check by this comparison. Large-scale properties can be estimated from many different clues: the density of the populations of galaxies, the magnitude and color of their light, the radio emissions signaling collisions and other significant events, the relation between the red-shift and distance of galaxies, and so forth.

Secondly, there are tests which can be made without looking so far away from home. We can think of the formation of new galaxies as equivalent to birth in the biological sense, and of their separation by expansion as equivalent to death. In terms of this analogy a new generation of galaxies is born, not every 30 years as in the case of human beings, but every few billion years. Now just as an animal population becomes extinct if it fails to reproduce its numbers from generation to generation, so large-scale properties of the systems of galaxies fail to survive unless they reproduce themselves in the same sense. If the universe is infinitely old, as the steady-state theory says, we should expect to see surviving only properties which have stabilized themselves so that they are reproduced at precisely the same level from generation to generation. In other words, according to the steady-state theory the galaxies are not a product of random fluctuations and condensations, as in the evolutionary theory, but represent a very strictly controlled system obeying a kind of cosmic ecology, with the origin of matter playing a critical role. This crucial difference between the two theories can form the basis of stringent tests. The tests can be applied to such properties as the density of galaxies in space and the distribution of sizes in masses of galaxies. That is, we can check whether the distribution follows a regular frequency curve or shows no regular pattern.

During the past five years it has twice been claimed that observations disproved the steady-state theory, but it now appears that in both cases the observations are open to serious doubt. The U. S. astronomers Joel Stebbins and A. E. Whitford thought that certain distant galaxies showed more reddening

than could be attributed to the usual red-shift, and this was construed to support the evolutionary theory. But Whitford later found that certain data they had made use of were incorrect. Recently Martin Ryle in England reported a count of radio sources which indicated that the density of galaxies in space increases with distance from us—again an apparent support for the evolutionary hypothesis. However, Ryle's findings have been questioned by the radio astronomer B. Y. Mills in Australia.

In my view the most serious potential contradiction of the steady-state theory lies in the recent red-shift studies by the astronomers in California. The findings so far, however, are highly uncertain.

George Gamow has offered against the steady-state theory the objection that elliptical galaxies (which are thought to consist only of old stars) apparently do not show the age variations that the theory predicts. In defense of the theory it can be said that the measurements cited (studies of the color of the galaxies, in two colors) are not a very sensitive index of the galaxies' ages. In the color test a galaxy six billion years old should look much like one three billion years old. More sensitive measurements are required.

The steady-state theory gains support, on the other hand, from recent studies indicating that the elements beyond hydrogen are formed in stars. These studies make it appear more likely that the elements are constantly being "cooked" in the stars, as the steady-state cosmology suggests, than that they were created in a primeval explosion, as Gamow has urged.

Radio astronomy offers the exciting possibility of something close to a direct test of the creation of matter in space. The total amount of matter in the galaxies we can observe is estimated to come to about 10^{-30} of a gram per cubic centimeter if it were spread evenly all through space. The steady-state theory predicts that the average density of matter should be 10 or more times greater than this. The difference, according to the theory, is accounted for by hydrogen spread through intergalactic space. Up to now it has not been possible to detect intergalactic matter. But in the next few years new radio telescopes, tuned to the one-note "song of hydrogen," may be able to test whether such quantities of hydrogen do or do not exist in space.

The Author

FRED HOYLE is lecturer in mathematics at St. John's College, University of Cambridge. He was born in Yorkshire in 1914; by the time he was six, he had taught himself the multiplication tables up to 12 times 12. Not all of his early experience with mathematics was so prodigious. One of his first teachers slapped him for miscounting the number of petals on a flower; his sense of justice was so outraged that he refused to return to the school. Failing eyesight caused him to give up cricket when he was 13, but did not prevent him from staying up all night with a three-inch telescope his parents had bought for him. Now a senior fellow at St. John's, Hoyle is widely known for his books *The Nature of the Universe* and *Frontiers of Astronomy*.

Bibliography

COSMOLOGY. H. Bondi. Cambridge University Press, 1952.

THE STEADY-STATE THEORY OF THE EXPANDING UNIVERSE. H. Bondi and T. Gold in *Monthly Notices of the Royal Astronomical Society*, Vol. 108, No. 3, pages 252-270; 1948.

SCIENTIFIC AMERICAN August 1953, Vol. 189, No. 2, pp. 23-29 OFFPRINT 219

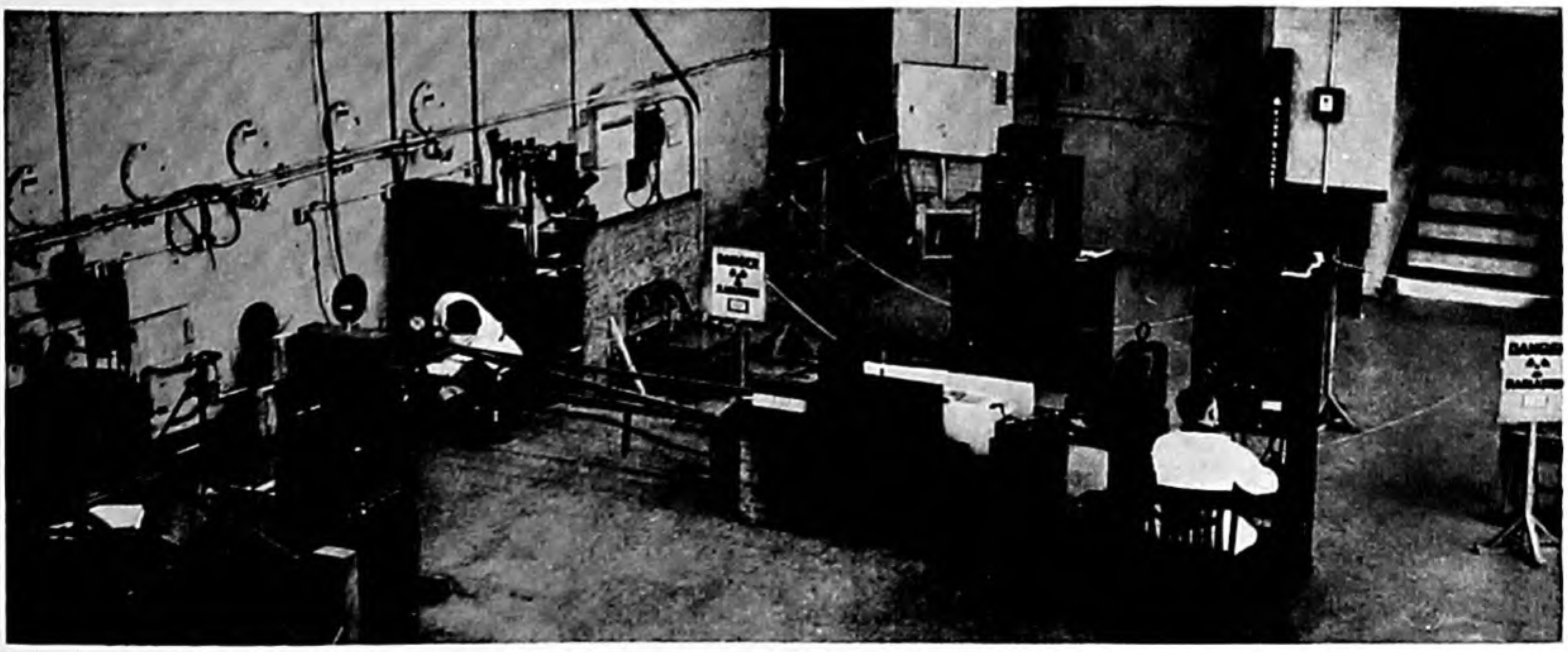
THE NUCLEAR REACTOR AS A RESEARCH INSTRUMENT

by Donald J. Hughes

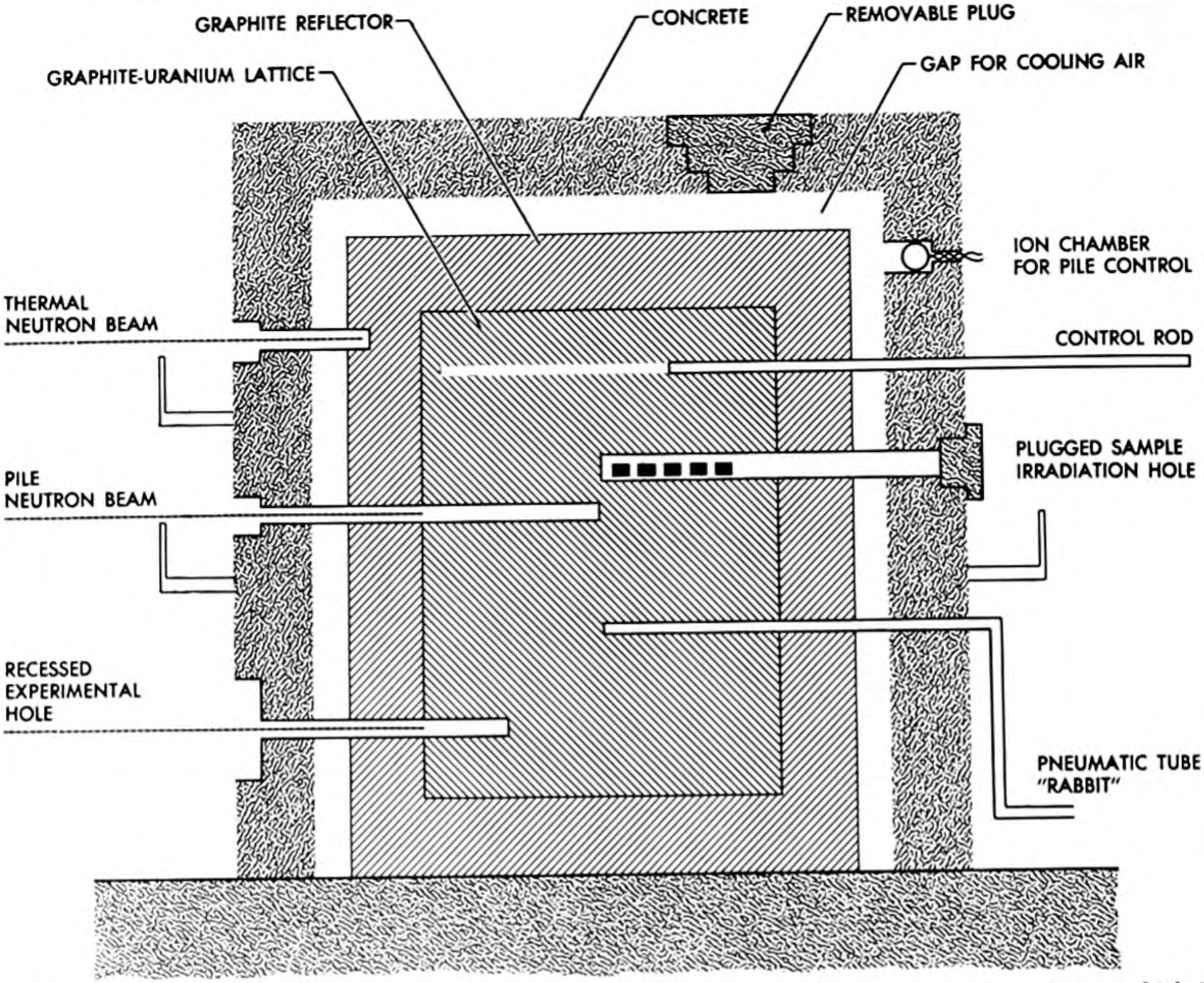
The enormous quantities of neutrons released by atomic chain reactions provide physicists with a potent means of studying matter. A survey of some of the ingenious techniques and results of pile neutron research.

NUCLEAR reactors manufacture, among other things, plutonium, heat and neutrons. Public attention has centered on the first product, as an ingredient of atom bombs, and the second, as a potential power source. Scientists are almost exclusively concerned with the third. As fundamental constituents of matter, neutrons are interesting in themselves, and they are one of the most powerful tools in the kit of the experimental physicist [see "The Neutron," by Philip and Emily Morrison; SCIENTIFIC AMERICAN Offprint 232]. In pre-reactor days we had to make neutrons by hand, so to speak, and in small quantities. With reactors we are mass-producing them in numbers beyond our wildest dreams of a few years ago. The torrent of neutrons now available provides us with an ideal means of investigating one of the central problems of physics—the nature of matter.

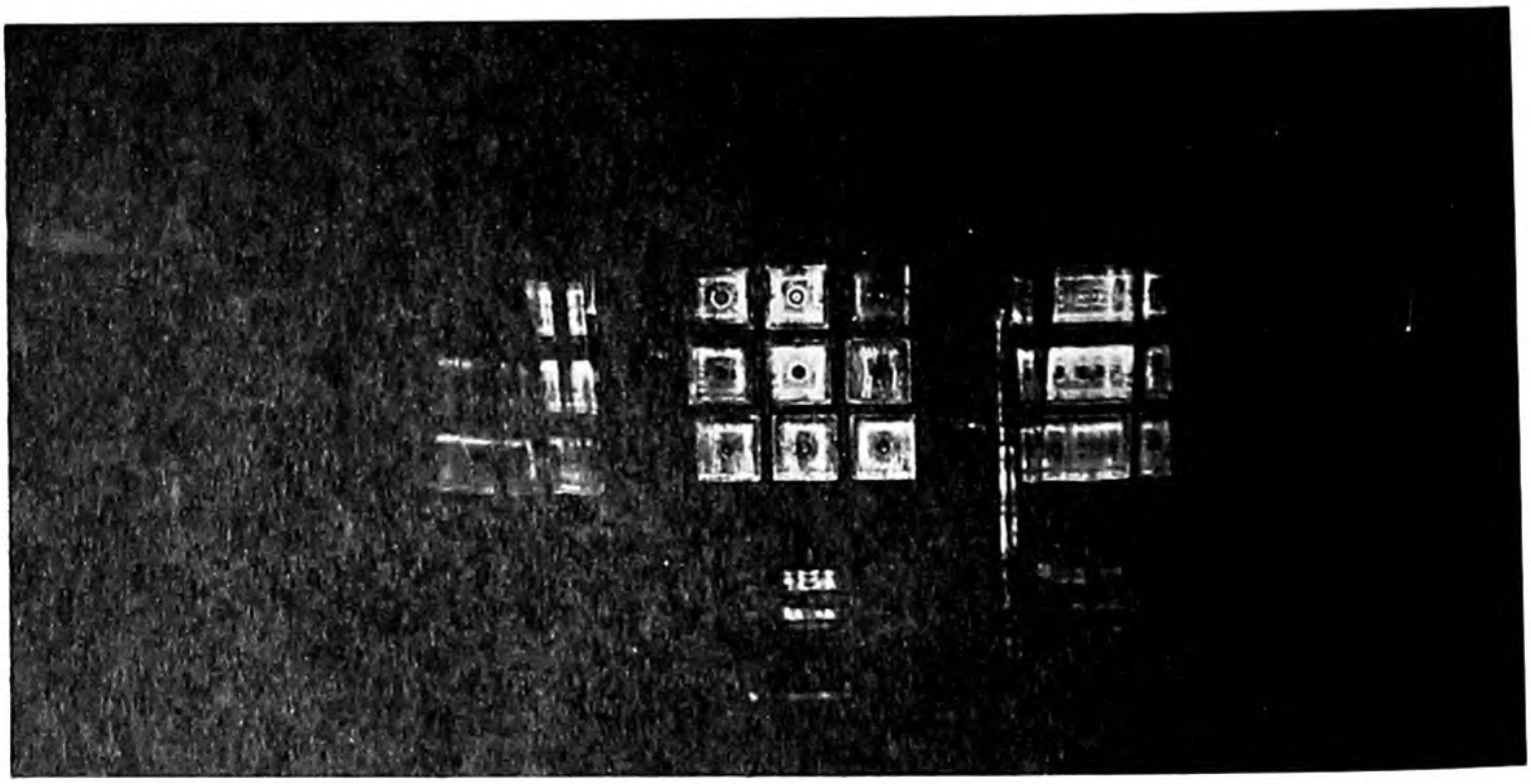
Why are neutrons so effective as probes of matter? First, they have no electric charge. Unaffected by the clouds of negative charge with which electrons surround atomic nuclei, neutrons pass easily through swarms of atoms. When they finally chance to strike a nucleus, neutrons penetrate it easily, being equally unaffected by its positive charge. Especially important is the fact that even very slow neutrons are highly penetrating. All other experimental projectiles, such as protons or alpha particles, must move so fast to get into the nucleus of the atom that they cannot be used to investigate the intricate details of its



REACTOR at Brookhaven National Laboratory is used solely for experimental purposes. The shield of the reactor is at the left; it is pierced with holes that emit neutrons. Here most of the holes are plugged, but one is open for a neutron-reflection experiment. The neutron "mirror" is at the left end of the long apparatus.



CROSS SECTION of the Brookhaven reactor shows its experimental features in simplified form. The uranium fuel of the reactor is embedded in a graphite moderator, which is surrounded by a graphite reflector, which is surrounded by a concrete shield. The moderator and reflector are pierced with holes that emit the neutrons.



HOLE extending into the center of the Brookhaven reactor is a foot square and 20 feet deep. At the far end of the hole may be seen an arrangement of bismuth blocks. The neutrons that emerge from this hole are utilized in experiments involving the "fast chopper" (*see illustrations at the bottom of the opposite page*).

structure. Such details can only be observed with slow neutrons of accurately controlled velocity.

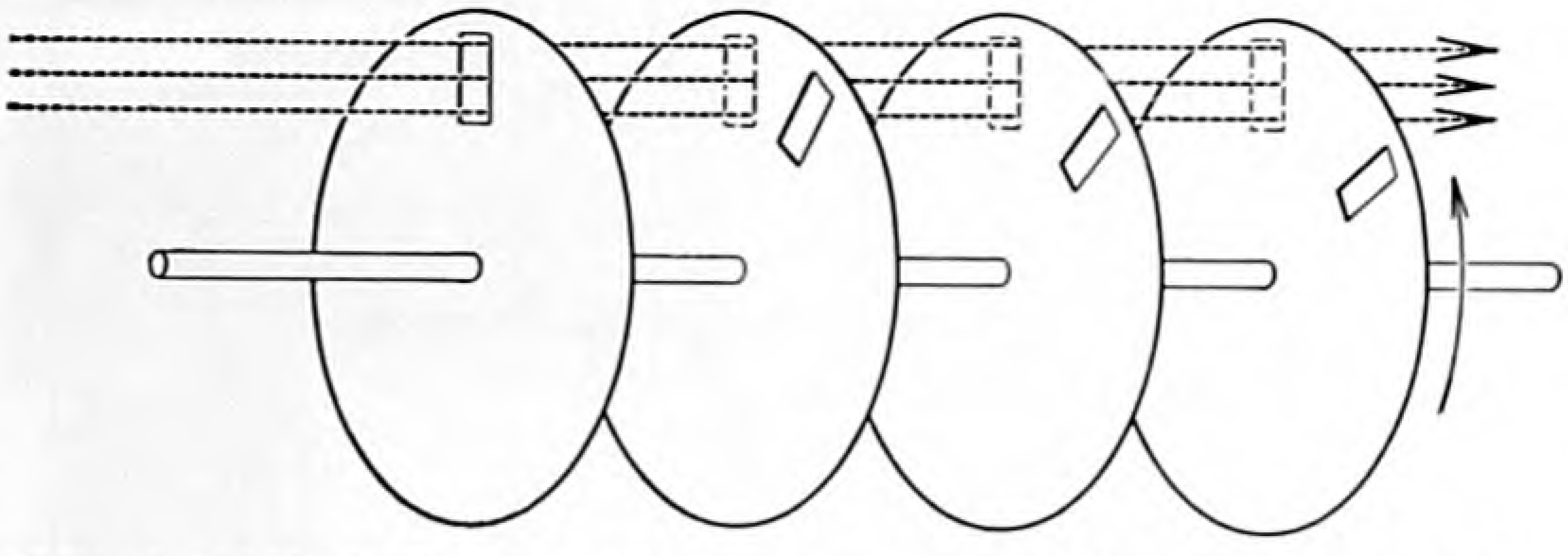
Second, the neutron is magnetized. Thus it reacts with the magnetic fields of atoms, yielding valuable information about magnetic properties. Finally, the neutron is a wave; or rather, like all bits of matter, it is both particle and wave. Some of the most informative experiments we can perform are based on the wavelike interaction of neutrons with matter. Only the intense beams that reactors provide make it possible to take practical advantage of the wave nature of neutrons.

THE very quality that makes neutrons so valuable as an experimental probe makes them hard to utilize. Charged particles are docile beasts. We can speed them up or slow them down with electric fields, steer and aim them with magnets. The uncharged neutrons, however, are unaffected by these forces and are difficult to maneuver. Even to know where they are is something of a problem since they do not announce themselves directly by leaving ion trails in cloud chambers or photographic emulsions as charged particles do. Before neutrons could be put to use a great deal of experimental ingenuity went into learning how to manipulate them.

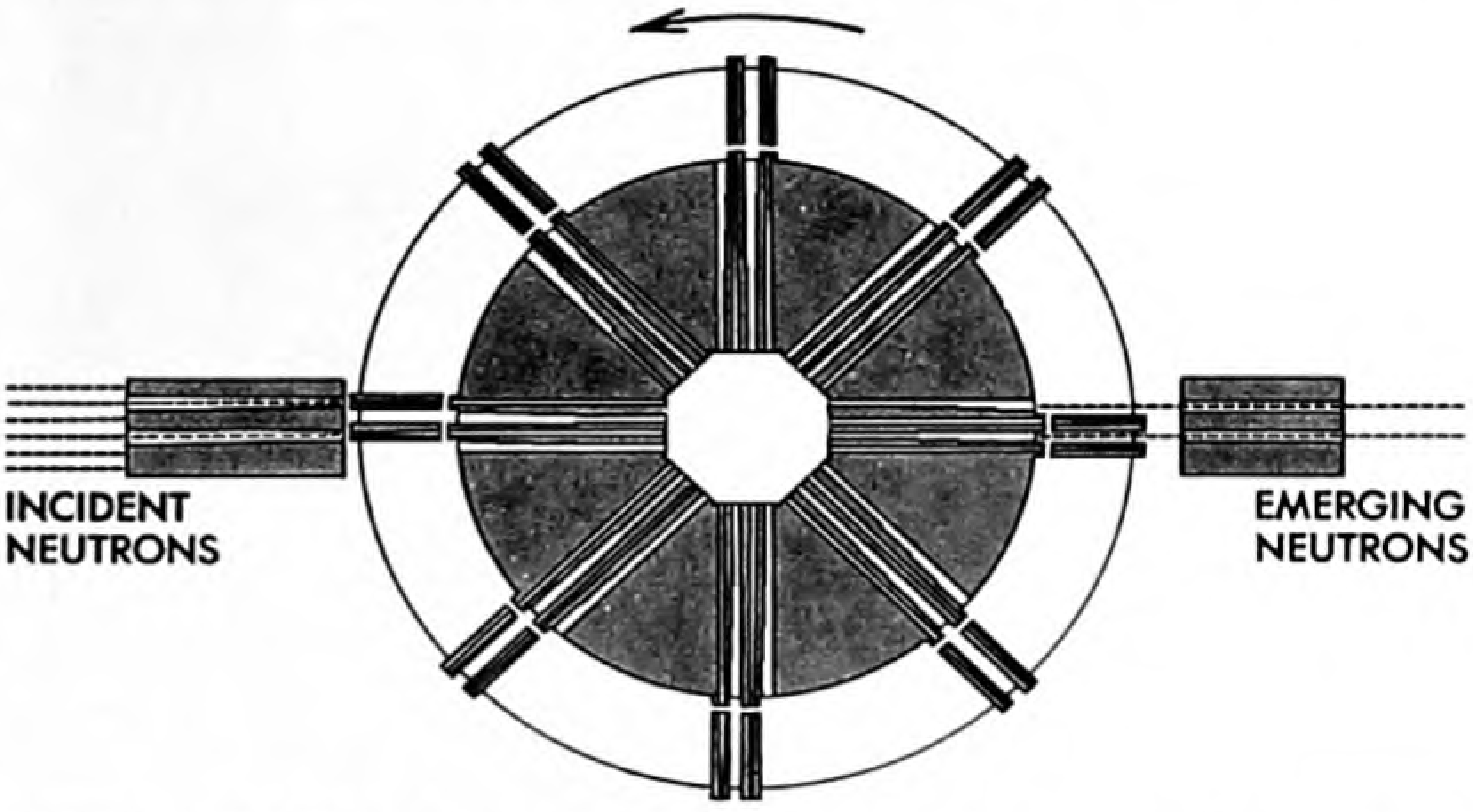
Experimenting with neutrons means, in essence, exposing a piece of material to them and observing how they are affected. Some may be absorbed, others deflected from their original path. The "cross section" of various atomic nuclei for absorbing or scattering neutrons (*i.e.*, the probability that the atom will absorb or scatter a neutron that hits it) tells us a good deal about the atoms and about the neutrons themselves.

A direct and comparatively simple way to use reactor neutrons for cross-section measurements is to put the material being studied inside the reactor. Despite their great size, reactors are extremely sensitive to changes in the number of neutrons available to maintain their chain reaction. An accurate measure of the availability of neutrons in the reactor can be obtained by noting how far the reactor's control rod must be pulled out in order to get the chain reaction started. When even a tiny sample of foreign, neutron-absorbing material is inserted in the reactor, the control rod will have to be pulled out measurably farther than normally. The extra distance measures the neutron absorption of the sample and hence its cross section. Another way to get this result is with the "pile oscillator." Here the sample is moved in and out of the reactor at a steady rate so that its power level rises and falls.

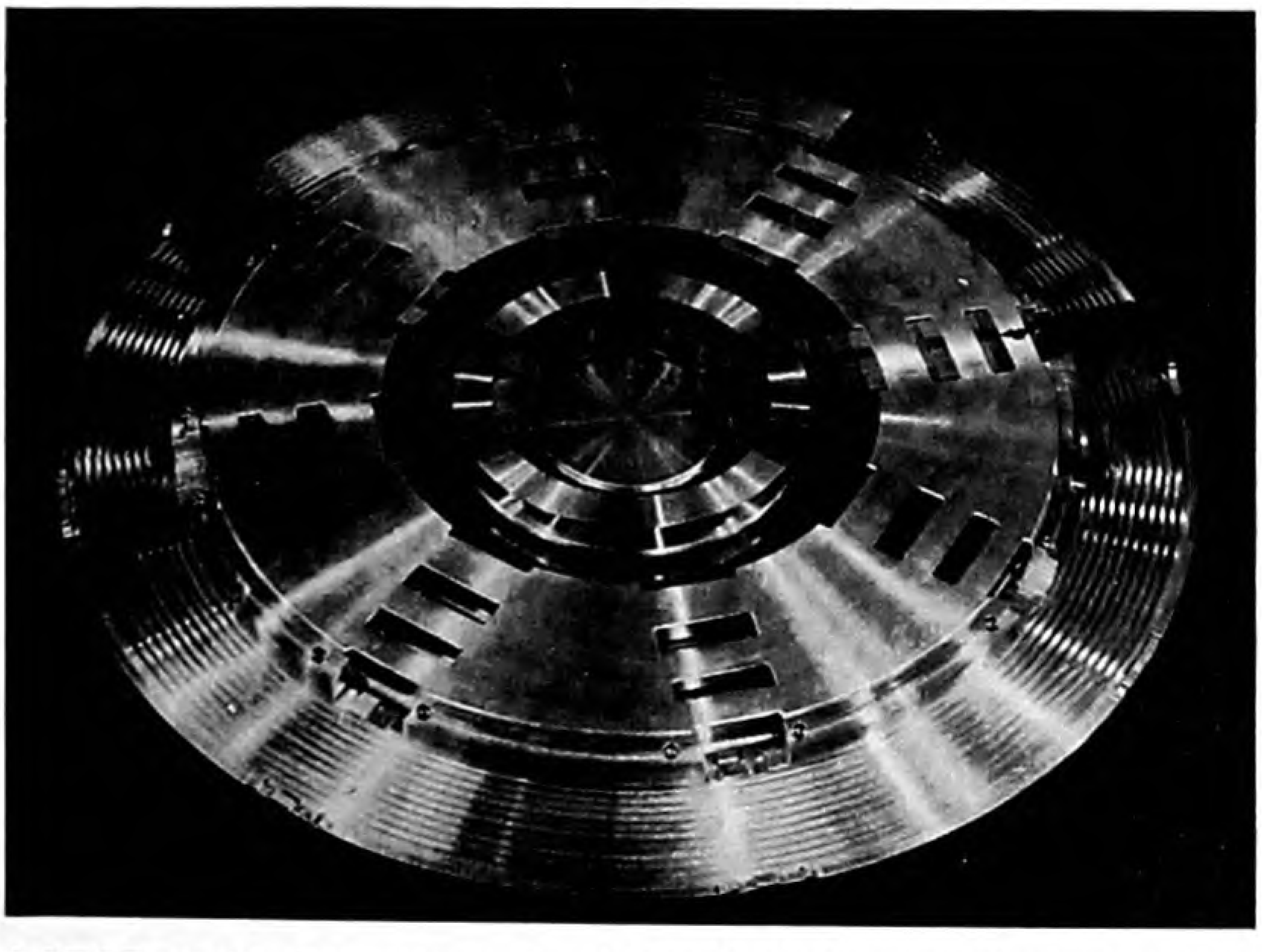
IN THESE experiments the reactor is itself a measuring instrument. For most work, however, it serves merely as



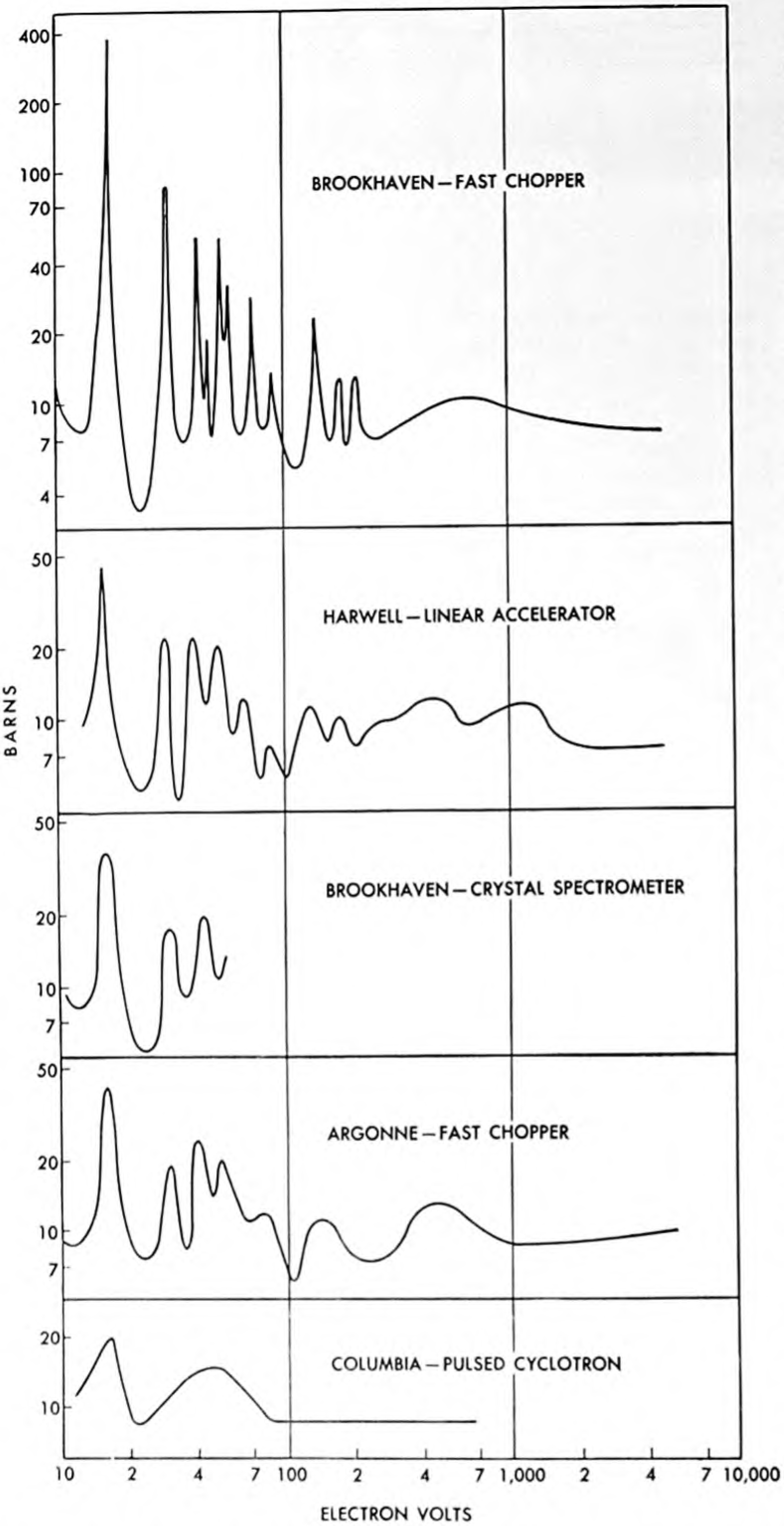
SPIRAL MONOCHROMATOR produces a beam of slow neutrons with uniform speed. Its slotted cadmium disks are turned on the same shaft. Only neutrons of a narrow range of speed will pass through all four of the slots.



FAST CHOPPER is shown in cross-sectional top view. It is a disk traversed by 16 pairs of channels through which neutrons can pass. When the disk is whirled in a vacuum at high speed, the neutrons emerge at right in bursts.



ROTOR of the chopper is made of forged aluminum with plastic inserts which form the channels through which the neutrons pass. When the rotor is spun at full speed, it is able to chop 1,000 bursts of neutrons per second.



HIGH RESOLUTION of the Brookhaven fast chopper in the measurement of nuclear "cross sections" is illustrated by these curves. The cross section of the nucleus, measured in "barns," is the probability that it will absorb a bombarding particle. Each curve represents the cross section of the silver nucleus when bombarded with neutrons of various energies, measured in electron volts. Because the Brookhaven chopper produces neutrons of precisely known energy, it is able to reveal cross sections in unusually fine detail.

a neutron source. Neutrons are brought out of the pile through holes running from the interior out through the shield. The beam that pours out of the experimental holes contains enormous numbers of particles in a wide range of energies. The fastest, coming directly from fissioning uranium atoms, have energies as high as 10 million electron volts. The slowest, those that have lost almost all of their energy in passing through the graphite (or, in the case of some reactors, heavy water) moderator, have an energy of only .0001 electron volt. Thus the fastest neutrons are some 100 billion times more energetic than the slowest. Each particle, it must be remembered, is also a wave. The formula for particle wavelength tells us that it varies inversely with speed. The fastest neutrons in the beam have wavelengths of a few hundred-thousandths of an angstrom unit, even shorter than the diameter of an atomic nucleus (an angstrom unit is a hundred-millionth of a centimeter). The slowest neutrons, on the other hand, have a wavelength of 30 angstrom units, far longer than the distances between atoms in crystals.

In different sections of this broad energy spectrum the neutrons have wholly different properties, both as to the way they must be manipulated and as to their effects on matter. High-energy neutrons behave like solid pellets; slow neutrons lose their particle character almost entirely and act like almost pure waves. The principal task of the reactor experimenter is to sort neutrons of various speeds and to observe how particles of each speed interact with matter.

THE NEUTRON beam divides into three main speed ranges. The so-called fast neutrons are those with energies extending from 10 million down to about 10,000 electron volts. Below these in the hierarchy of energies are "resonance" neutrons, from 10,000 electron volts to .01 electron volt. Finally there are the thermal neutrons, with energies from .01 to .0001 electron volt.

In many experiments we do not try physically to separate the various particles from one another. Instead we use detecting devices that respond only to the type of neutron we are studying. Thus fast neutrons make themselves known by knocking protons out of a detecting substance such as paraffin. In cloud chambers the protons produce ion trails from which we can deduce the energy of the particles, and hence the energy of the neutrons that produced them. Counting the proton tracks of a particular energy tells us how many neutrons of corresponding energy are in the beam. Suppose we want to measure the cross section of gold for million-volt neutrons. First we expose the detector to the reactor beam and count only the million-volt neutrons. Then we put a

thin gold foil into the beam and note the decrease in these particles, which represents the number of million-volt neutrons removed by the gold. Most basic research in the fast range deals with nuclear cross sections, either for absorption or scattering.

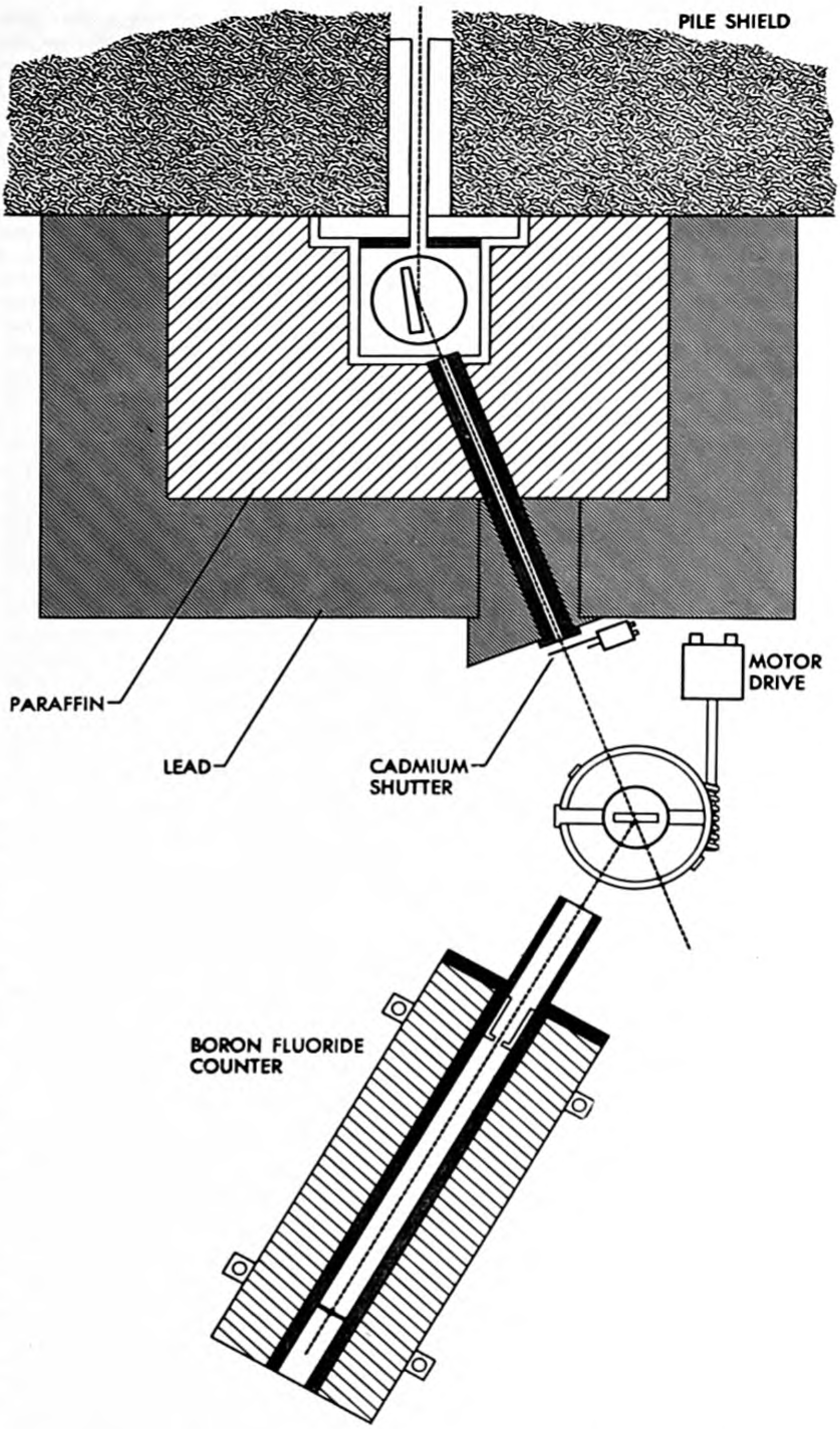
RESONANCE neutrons are a kind of mixed breed. The faster ones must be handled like particles, but their effects depend on their wave nature. Having wavelengths of about the same size as nuclear diameters, they interact with nuclei in a wavelike manner. At certain wavelengths they resonate with a nucleus, which means they are strongly absorbed; at slightly different wavelengths they are "out of tune" and their absorption greatly decreases. The neutron absorption curve of a particular nucleus thus shows sharply separated peaks at isolated wavelengths.

Nuclear physicists are very interested in these curves because they promise to throw light on one of the murkiest problems now facing us—the internal structure of the nucleus [see "The Structure of the Nucleus," by Maria G. Mayer; SCIENTIFIC AMERICAN Offprint 228]. Is it formless, like a liquid drop, or does it have a "shell" structure? The shell theory states that neutrons and protons occupy fixed orbits or shells in the nucleus, much as the electrons do outside it. A nucleus whose outermost shell is completely filled should be very stable. The numbers of neutrons and protons required to fill successive shells are known as the "magic numbers."

Neutron absorption experiments give some strong evidence for the shell model. One of the magic numbers is 50, and tin contains 50 protons. When we compare the absorption spectrum of tin with that of silver, we can see that tin is indeed unreactive. Whereas the silver nucleus finds many neutron energies that excite it, tin is apparently satisfied with its make-up and rejects all neutrons indifferently.

BUT we are still far from having a satisfactory theory of the nucleus. We need much more information on the activity and energy levels of various nuclei, and neutron absorption experiments are one of the best ways to get it. It is essential to find out just which energies are absorbed by each nucleus, and which rejected. This means that we must be able to distinguish sharply between neutrons of different speeds. The sharper our separation, the greater the resolving power of our nuclear spectroscope, and the less the chance that what appears to be a single absorption peak is really two or more separate ones.

The device with which we are now getting the sharpest separations is the "fast chopper," recently developed by F. G. Seidl at the Brookhaven National



NEUTRON DIFFRACTION experiments at Oak Ridge National Laboratory employ the arrangement depicted in this top view. The neutrons emerge from the reactor at the top. They are reflected from a crystal set at a shallow angle as a means of selecting those of a certain energy range. The neutrons are then allowed to pass through a second crystal mounted on a turntable. This crystal diffracts the neutrons in certain preferential directions; when it is turned, the counter at the bottom will detect the number of neutrons diffracted in each direction. In this way the experimenter can measure the interaction between the neutrons and the atoms in the crystal.

Laboratory. It is a solid disk traversed by a few narrow slits through which neutrons can pass. When the disk is whirled very rapidly it acts as a shutter for the beam, allowing short bursts of neutrons to pass each time a slit swings into the proper position. At full speed the chopper yields 1,000 pulses per second, each lasting a millionth of a second. The slits are an inch high by .01 inch wide. The extremely narrow beam that they produce makes it possible to work with small samples of material, as little as a hundredth of a gram. This means that we can experiment with separated isotopes, which are available only in fractions of a gram.

To detect the neutrons we use a crystal that scintillates when they strike it. The crystal is set up about 100 feet from the chopper. As each burst of neutrons traverses the long path, it becomes elongated in the direction of motion, the fast particles heading the parade and the slow ones bringing up the rear. Attached to the scintillating crystal are 100 separate counters, each turned on for successive intervals lasting a hundredth of the total flight time. A beam of light, which passes through the slits with each burst of neutrons, triggers the counting sequence, starting the first counter just in time to pick up the first neutrons. Thus each counter records particles in a very narrow range of speed. To record an absorption spectrum the counters are read first with an unobstructed neutron beam. Then a sample of material is placed between the chopper and detector and the counters read again. The silver spectra reproduced on page 152 show how much sharper resolution we get with the Brookhaven chopper than with earlier methods of selecting neutron velocities.

THE fast chopper works well with neutrons of about 10,000 to 10 electron volts. At the lower end of the resonance range, from 10 to .01 electron volt, the neutron waves are long enough for us to take advantage of their optical properties. These waves are about as long as the distance between atoms in crystal lattices and are diffracted by crystals just as X-rays are. The angle at which they are scattered in a given crystal depends on their wavelength, so that these neutrons can actually be separated according to speed in a crystal "monochromator."

Having obtained a beam of single-energy neutrons, we generally use it to study other crystals. X-rays, of course, have been used in the same way for half a century, but neutrons can do things that X-rays cannot. Neutron diffraction has opened an entirely new field in crystal analysis—the investigation of magnetic structure.

Every neutron is a tiny magnet. When it is scattered by magnetized atoms its deflection depends not only on the "grating spacing" of the crystal lattice but also on the magnetic forces. Neutron diffraction patterns are, therefore, different from X-ray patterns for magnetic substances, and from the difference we can read the magnetic configuration of the crystal.

At several laboratories neutron diffraction has been applied to the study of anti-ferromagnetic materials. In a ferromagnetic substance such as iron each atom is a magnet and all the atoms are lined up with their north poles facing in the same direction. Anti-ferromagnets also have magnetized atoms, but show no over-all magnetism. This has been explained by assuming that the individual atoms are aligned, but with their north poles pointing alternately in opposite directions. There was no way of checking this assumption, however, until C. G. Shull and J. S. Smart of the Oak Ridge National Laboratory tried neutron diffraction. They exposed a manganese oxide crystal to a monochromatic beam of neutrons. If the atoms are actually arranged according to the theory, a neutron will be attracted by one, repelled by the next, attracted by the third and so on. Because of this alternate attraction and repulsion the "unit cell"—the fundamental structural unit of the crystal—is twice as large when viewed by neutrons as by X-rays. The diffraction pattern should then contain neutrons at positions where there would be no energy in an X-ray pattern. Neutrons were found in the diffraction pattern in just these expected positions, but only when the crystal was kept at very low temperature. At room temperature the pattern disappeared, showing that the rigid anti-parallel alignment of the atoms no longer held good. Other anti-ferromagnetic substances are believed to have much more complicated magnetic structure than the simple scheme of manganese oxide. Knowing how their atomic magnets are arranged is important both for our basic understanding of the solid state and for a number of potentially useful applications. With neutron diffraction we can now find out.

BELOW the resonance range are the slowest neutrons, the thermals. They have this name because, by repeated collisions with carbon nuclei in the reactor, they have been slowed to speeds equal to, and even less than, the average vibrational speeds of the graphite molecules. These neutrons have energies from .01 to .0001 electron volt. At their slowest they travel only about 600 feet per second, corresponding to a temperature of only two degrees centigrade above absolute zero.

We have various ways to sort out thermal neutrons of different speeds. One of these is a slow chopper, which operates in the same way as the fast chopper. Another is the "spiral monochromator," an elaboration of the chopper idea that actually filters particles of a single speed out of the beam. It is made of a series of slotted disks on a single shaft, the position of the slots being adjustable (*see diagram on page 151*). The disks are set so that each successive slot stands at a fixed angle to the one before it, or, in other words, so that the open path through the array of disks is a spiral. When the apparatus is rotated, a particle entering the first slot at the right speed will find each following slot moving into position just in time to let it through. Particles at other speeds will bump into the solid disks. By adjusting the angle between the slots or the speed of rotation of the shaft we can pick out neutrons of any desired speed.

Very slow neutrons are particularly useful for studying the motions of atoms in crystals. Moving in the same speed range as the particles they bump into, the neutrons can gain or lose a large proportion of their energy in collisions. From the energies of the scattered neutrons we can infer the motions of the atoms in the crystal lattice.

THE WAVELENGTHS of thermal neutrons are so long that it becomes just possible to reflect them with mirrors. Mirror reflection requires a smooth surface, and smoothness is a relative matter which depends upon the wavelength involved. The longest neutron waves are still 1,000 times shorter than waves of visible light, so that an "optically smooth" surface still looks very rough to a neutron. But just as a highway can reflect light that strikes it at a grazing angle, so a mirror that is rough to neutrons can reflect them when they strike at an exceedingly small angle, about a tenth of a degree.

Neutron mirrors are an effective means of measuring nuclear forces. This is because the force between the neutron and the atomic nuclei of mirror materials determines the maximum glancing angle at which reflection is possible. We can measure this "critical angle" quite precisely, and so calculate the nuclear force accurately. The method can be applied to a variety of atoms, since many materials make suitable mirrors. We have used highly polished solids, such as iron, nickel and beryllium; liquids, such as mercury, water and benzene, whose surfaces are sufficiently flat when quiet; gases, such as helium and nitrogen, contained in vessels with polished walls.

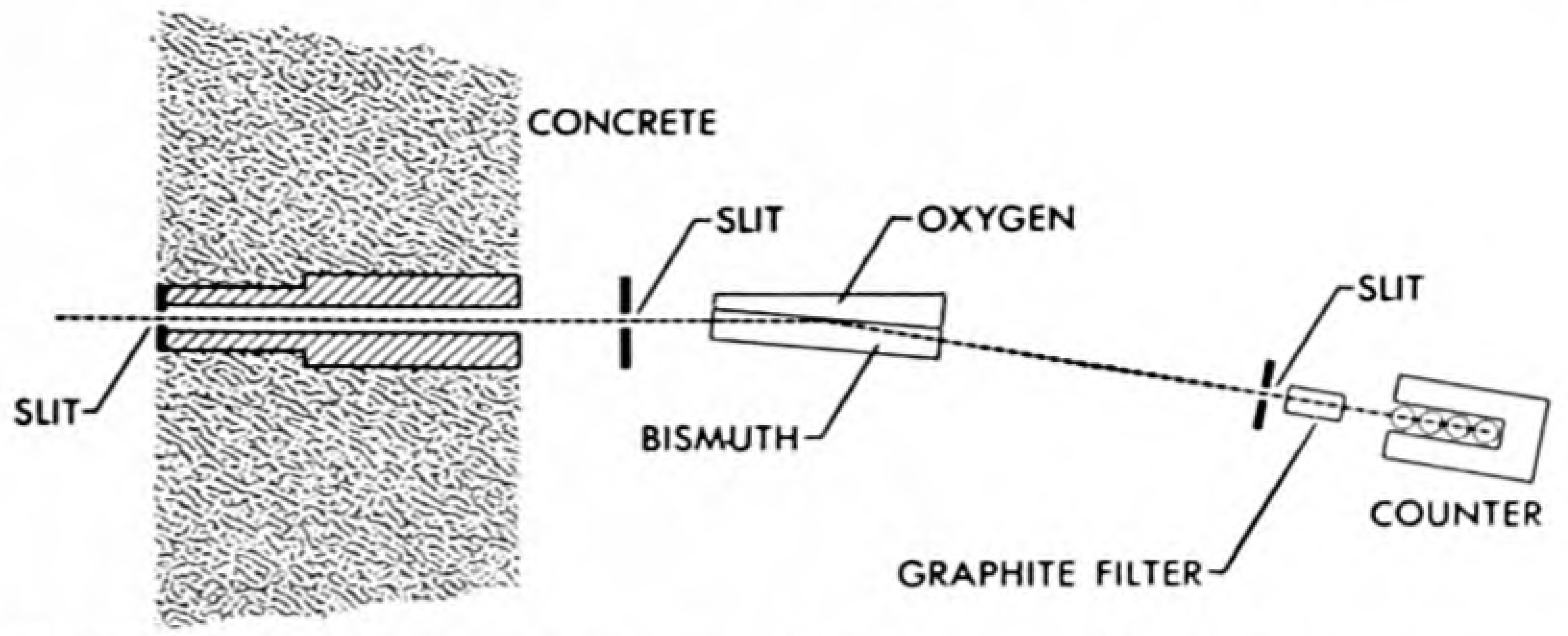
RECENTLY neutron mirrors have given us an accurate measurement of one of the smallest of all interactions between fundamental particles, the one between the neutron and the electron. It might seem at first that the force between these two fundamental particles should be zero, for the neutron has no

charge. According to modern meson theory, however, the neutron spends part of its life in a dissociated form, split into a proton surrounded by a cloud of mesons. It is still electrically neutral when seen from outside the meson cloud, because the negative charge of the mesons just balances the positive charge of the proton. But inside the cloud, which extends about 10^{-13} centimeter from the proton, there is an electric field. An electron inside the meson cloud should be attracted to the proton with a minute but measurable force.

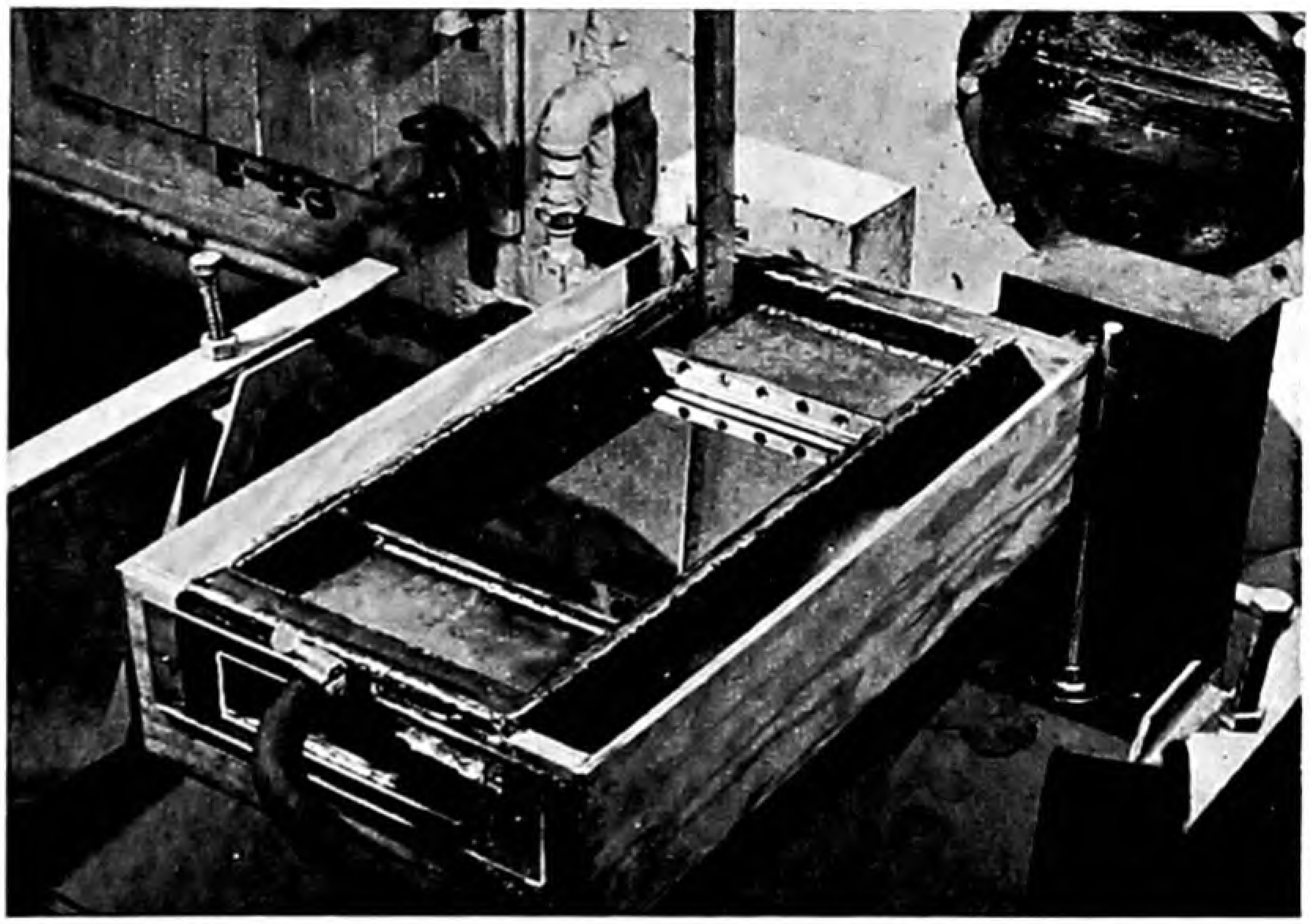
When we try to measure this force, however, we run into a major difficulty. The neutron-electron interaction is about 10,000 times weaker than that between the neutron and the atomic nucleus. It is impossible to study the interaction of neutrons with isolated electrons, and so the force we are after is always masked by one very much greater. To sidestep the large nuclear force the writer and his colleagues at Brookhaven, J. A. Harvey, M. D. Goldberg and Marilyn J. Stafne, devised a special neutron mirror. A block of bismuth was polished on one side and fitted into a tank so that a layer of liquid oxygen could be placed in contact with the polished surface. The neutrons were reflected from the boundary between the two materials. It happens that oxygen and bismuth have almost exactly the same nuclear interaction with neutrons, so the nuclear scattering balances for the two. The electron scattering, however, is quite different on opposite sides of the dividing boundary because each bismuth atom has 83 electrons and each oxygen atom only eight. So far as the neutrons are concerned the experiment is essentially the same as if they were reflected from a mirror made of electrons alone. The critical angle for the mirror thus gives the strength of the neutron-electron interaction directly.

The measured force turns out to be much smaller than predicted by the meson theory; in fact, it can be completely accounted for by a small magnetic interaction between the neutron and electron that does not involve the meson cloud at all. This means that there is some error in the present simple idea that neutrons exist transiently as protons and mesons. It may be that the proton itself, instead of being a point particle, is spread out to such an extent that its more distant parts act weakly on the electron; it may be that the splitting produces both positive and negative mesons, which cancel the electrical effect. To answer these questions fully we will have to design still subtler experiments as well as perhaps construct better theories of fundamental particles.

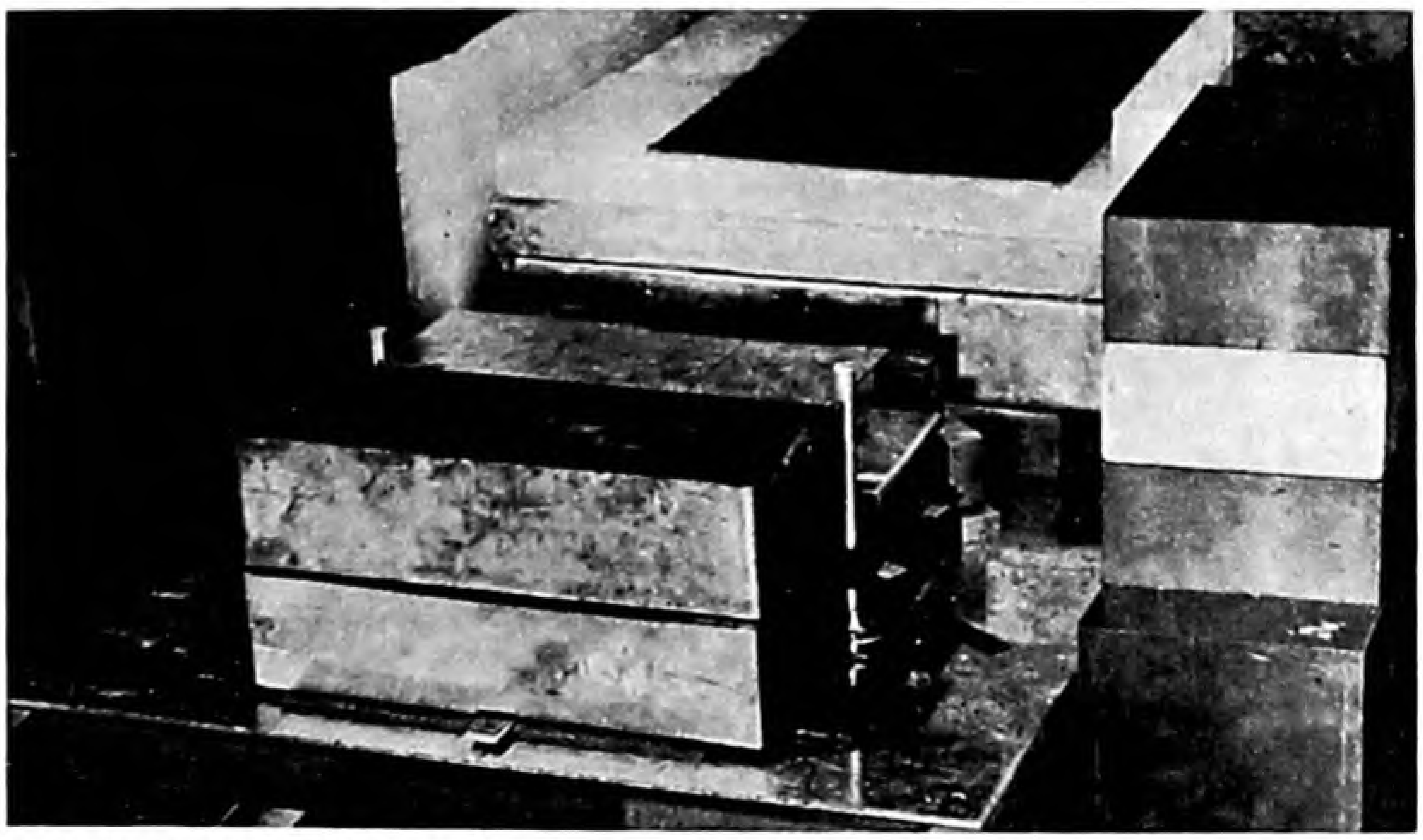
One thing seems certain. As more research reactors are built, and with higher neutron fluxes, their neutrons will play an increasingly important role in the search for the secrets of matter.



NEUTRON MIRROR experiments at Brookhaven utilized the arrangement shown in this side view. The neutrons emerged from the pile at the left and were reflected from the boundary between bismuth and liquid oxygen.



BISMUTH MIRROR is used in the arrangement depicted in the diagram above. During an experiment liquid oxygen is poured into the container around it. The neutrons emerge from the pile hole through the slit at right.



CADMIUM BLOCKS form a slit between the bismuth-oxygen mirror and the counter which detects neutrons reflected from it. The cadmium absorbs stray neutrons, as do the paraffin blocks that are seen in the background.

The Author

DONALD J. HUGHES has been working with atomic piles since the first one went into operation in a University of Chicago squash court. He was born in Chicago in 1915 and educated at the University of Chicago, receiving his Ph.D. in physics in 1940. Until early in 1943 Hughes worked at the Naval Ordnance Laboratory on mine and torpedo detectors. Then he transferred to the Manhattan District in Chicago and has been working in the field of nuclear physics ever since. He did research on pile neutrons first in Chicago with the original pile and later at the Hanford Engineer Works when the big plutonium-producing reactors were built. After the war he joined the Argonne National Laboratory as director of its nuclear physics division. In 1949 Hughes went to the Brookhaven National Laboratory, where he is now.

Bibliography

PILE NEUTRON RESEARCH TECHNIQUES—I. D. J. Hughes in *Nucleonics*, Vol. 6, No. 2, pages 5-17; February, 1950.

PILE NEUTRON RESEARCH TECHNIQUES—II. D. J. Hughes in *Nucleonics*, Vol. 6, No. 5, pages 38-53; May, 1950.

PILE NEUTRON RESEARCH TECHNIQUES—III. D. J. Hughes, Vol. 6, No. 6, pages 50-55; June, 1950.

SCIENTIFIC AMERICAN August 1949, Vol. 181, No. 2, pp. 48-51 OFFPRINT 220

RADIOACTIVITY AND TIME

by P. M. Hurley

Half a dozen or more decaying elements trapped in the rocks now serve as clocks by which physicists can date events in the history of the earth going back to billions of years ago.

ABOUT 15 years ago Eric Temple Bell, professor of mathematics at the California Institute of Technology, wrote a science-fiction tale called "Before the Dawn." A piquant item in this fantasy was an electronic televisor that projected events not in space but in time. When focused on ancient rocks, it could reconstruct on a large screen a motion picture of the story that the rocks had to tell. The machine re-enacted with photographic accuracy geological upheavals that had occurred on the site eons ago, and battles between animals now long extinct.

If modern geology cannot match the realism of Bell's automatic rock-reader, it can reconstruct the history of the earth with some accuracy and detail, at least for the last 15 per cent of the earth's age, the period covered by the fossil record. And it now has a technique which seems even more unreal than Bell's machine—a natural clock that gives us the time of events in the dim geologic past. In other words, for the first time we can read in the rocks not only the sequence of happenings but their dates.

The discovery that made this possible was radioactivity. The slow decay of naturally radioactive elements in the earth's crust has ticked off the seconds since the earth was born with unalterable precision. The rates of disintegration of these elements are so unaffected by the temperatures and pressures of our planet that they may be considered fixed constants. Thus by measuring the amount of the elements' decay we can determine the age of the structure of which they are a part.

Effectively, the clocks began to tick when some molten part of the earth solidified into crystalline rock. When a single mineral crystal is formed, it contains certain predominating elements and a number of minor ones trapped in its structure as impurities. The radioactive element that serves us as a clock may be either a major or a minor constituent of this crystal. Let us say it is uranium. The element breaks down at a known rate, producing certain isotopes of lead as stable end-products. After a few tens of millions of years there is enough such lead in a sample of the rock to be measured chemically. By measuring the end product and the amount of radioactive uranium still left, one can determine the age of that particular rock. This in turn dates an event associated with the rock in the earth's history.

The method can be made clearer by considering a study of an actual mineral sample. The sample was a piece of samarskite, a velvety black, crystalline mineral, taken from the Spinelli Quarry in Glastonbury, Conn. This mineral contains uranium and thorium, both of which were utilized as clocks. In every 100 grams of the sample there were about seven grams of uranium, three grams of thorium and slightly more than three tenths of a gram of lead.

How much of this lead was the product of the breakdown of uranium and thorium, and how much of it was originally present in the samarskite when it crystallized? To find out, Alfred O. Nier of the University of Minnesota separated the tiny quantity of lead into its isotopes and measured them by means of a mass spectrometer. He measured the amounts of each of the four isotopes present: lead 204, 206, 207 and 208. Lead 206 is known to be a breakdown product, or daughter, of uranium 238; lead 207 is a daughter of uranium 235, and lead 208 is a daughter of thorium 232. But lead 204, fortunately, is not a daughter; it is a stable material, the amount of which has not been significantly increased by the breakdown of any possible progenitor in the past two or three billion years. Thus it serves as a measure of the amount of lead originally present in the sample.

In common lead the isotopes 204, 206, 207 and 208 appear in certain almost constant proportions. Assuming that the isotopes were present in these proportions in the samarskite before the breakdown products were added, the additions can be computed by measuring the change in the relative abundance of each isotope as compared with 204. The samarskite sample had .0004 of a gram of 204 in every 100 grams of the mineral. By comparing the amounts of the other isotopes in this sample with the amounts of the same isotopes associated with every .0004 of a gram of lead 204 in common lead, the amounts of added or radiogenic lead were determined.

Consider first thorium and its daughter, lead 208. In the samarskite sample there was .0518 of a gram of 208 for each .0004 of a gram of 204. In common lead the proportion is .0142 of a gram of 208 to .0004 of a gram of 204. Thus the difference, .0376 of a gram, was the amount of lead 208 produced by the disintegration of the original thorium. There were three grams of thorium left in the sample; from this the original amount of thorium could be computed. The half-life of thorium, *i.e.*, the time it would take for half of its atoms to disintegrate, is known to be 13.9 billion years. On the basis of these figures, the thorium clock showed that this rock was 266 million years old.

NO SINGLE clock is entirely trustworthy, however, in this type of investigation. There is no assurance that we have an accurate estimate of the amount of the parent isotope originally present, or that some of the parent or daughter atoms have not escaped from the rock. For example, the crystals have been under constant bombardment for millions of years by alpha particles emitted by the radioactive elements. This barrage shatters some of the crystals, and some atoms may well leak out of the rock, particularly in the water-saturated zone near the earth's surface. In that case the amount of disintegration of the parent element cannot be determined accurately.

In the particular case of the samarskite sample, fortunately, there were no fewer than three other clocks available for checking. One was uranium 238, which has a half-life of 4.5 billion years and decays to lead 206 as its end product. The amount of its daughter lead was computed to be .236 of a gram per 100 grams, and so this clock gave 255 million years as the age of the rock. The second clock was uranium 235, which has a half-life of 707 million years and

yields lead 207. It gave 254 million years as the age of the rock. The third clock was based on the proportion of lead 207 to lead 206; because the half-lives of their parents, U-235 and U-238, differ, this ratio varies with time in a calculable manner. This clock timed the age of the rock as 256 million years. Thus the three uranium clocks showed remarkable agreement. The thorium measurement, apparently less accurate, was discarded, and the age of the sample was considered to be accurately fixed at about 255 million years.

Such close agreement is, of course, the exception rather than the rule. Moreover, it is not often that more than one clock is at hand to check the results. Because minerals with enough uranium or thorium in them to be usable for lead ratios are relatively rare, age determinations by this method frequently cannot be made where needed most. What geologists have sought is a measure that can be used on the small amount of radioactivity present in common rocks. Virtually all rocks have at least a little radioactivity; granites, for example, average about a tenth of an ounce of uranium per ton.

There was great hope for a time that helium would provide this sensitive measure. Helium, whose atomic nuclei are alpha particles, is one of the products of the breakdown of uranium and thorium; *e.g.*, each atom of U-238, in decaying to stable lead, yields eight atoms of helium. Measurements of helium in rock samples at first seemed to give rock ages that agreed with those determined by lead measurements, but it was found later that this agreement was accidental and due to a compensating error.

There is almost always less helium in rocks that have crystallized out of a molten state than one would expect. In such rocks radioactive elements are concentrated in little pockets. Helium in the

AGES OF ROCKS in North America have been outlined by measurements based on radioactivity. Black and white strips indicate the roots of mountain belts, some of which are numbered in their order of development. Dotted lines denote possible mountain belts. Type of radioactive measurement is given by the ratio of one element to another, *e.g.*, a lead-uranium ratio is indicated as Pb/U. Rock ages are given in millions of years.

form of alpha particles is fired atom by atom into the crystal structures of the rock as bullets might be fired into a log. It is believed that the crystal structures eventually break down and allow some helium to escape from the rock. Rocks that have not crystallized out of a molten state do not bear their uranium and thorium in these highly concentrated pockets, so they are being investigated for their usefulness in the measurement of time by helium ratios. One mineral, magnetite, has been tested extensively, and the results derived from it are in general agreement with those derived from lead ratios.

The element rubidium is another useful clock. The radioactive isotope rubidium 87, with a half-life of about 58 billion years, decays to a stable isotope of strontium. Thus the age of minerals containing rubidium can be found by determining the ratio of Sr^{87} to Rb^{87}. Certain micas are rich in rubidium. In lithium mica it has been found that more than 98 per cent of the strontium present is the product of the decay of rubidium. Measurements of a number of minerals containing rubidium have yielded ages that agree reasonably well with those based on uranium and lead.

These various clocks have now given us an approximately accurate time scale for the geological periods since the beginning of the Cambrian Period of some 500 million years ago (*see chart on page 160*]. In the map on page 158 are calculated ages of some older rocks in North America. J. Tuzo Wilson of the University of Toronto has recently plotted belts of exposed pre-Cambrian rocks in the ancient formation known as the Canadian shield. From observations at places where these belts crossed one another, he determined their relative ages. The radioactive clocks agreed in general with Wilson's evidence. For instance, rocks in the belt that he judged to be the oldest showed the greatest age, two billion years, in the radioactivity measurements.

BUT WHAT about dating events in the Pleistocene—the brief last million years in which man has lived on the earth? The slow-paced clocks that we have been considering have no minute-hands; they are not refined enough to measure intervals as short as a fraction of a million years, the period in which anthropologists and glacial geologists are particularly interested. The events of this period have hitherto been dated by the ebb and flow of glaciers, by tree rings and by sedimentary layers—all of which have been unsatisfactorily restricted as clocks.

To measure lengths of time of a few thousand or hundred thousand years, we must use isotopes with half-lives of the same order of magnitude, instead of the millions or billions of years of uranium and thorium. It so happens that there are several convenient ones.

One useful clock is found in the sediments laid down on ocean floors. In their microorganic remains the sediments contain an unbroken record of changes in climatic conditions and surface-water temperatures over the whole Pleistocene Period. And mixed with the sediments, deposited in them out of sea water, is the radioactive isotope thorium 230, with a half-life of about 83,000 years. Thorium 230 is one of the intermediate products in the decay of uranium to lead. It has been found that the ratio of thorium 230 to uranium decreases in a regular manner with increasing depth of the sediments. If the amounts of thorium 230 and uranium are measured in closely spaced samples taken from the top of the sediments down, the rate of deposition of the thorium at any level, and the age of that level, can be calculated.

It is now possible to obtain samples of these sediments from any desired depth and location by means of a ship-borne rig that cuts a long core out of the ocean bottom. Cores 20 to 30 feet in length, representing the entire Pleistocene or even longer, have been brought up and studied. One interesting and important investigation in the Antarctic region has shown that glacial conditions have existed there for 1.1 million years, and that ice ages in the Southern Hemisphere apparently were contemporaneous with those in the Northern Hemisphere.

This clock will be particularly helpful to anthropologists, who are interested in the dates of glaciation and climatic change. The ethnologist, unearthing an early culture of man, needs a shorter time measurement. A clock suited to his needs may be forthcoming as a result of a most happy recent discovery by E. C. Anderson, W. F. Libby and others at the University of Chicago. This clock is the radioactive isotope carbon 14. Anderson and his co-workers found that carbon 14 is being created continuously in the atmosphere by cosmic radiation. Neutrons produced by cosmic rays convert part of the nitrogen 14 in the air into carbon 14. The amount of carbon 14 is not continuously increasing, but has reached a balance point where as much breaks down by radioactive decay as is formed. Thus the small amount of carbon 14 in the air remains constant, and it was probably the same hundreds of thousands of years ago as it is now. But any carbon 14 that is taken out of the air by some stable material on the earth (*e.g.*, a tree or animal) decays without replenishment if it is buried. It is possible, therefore, to date wood, sea shells or other materials containing carbon that are found with artifacts in a buried camp site or cave simply by measuring the proportion of the carbon 14 that is left in the buried remains. Since the half-life of carbon 14 is only 5,700 years, this clock will be reasonably accurate only up to about 25,000 years. Recently the Chicago investigators made an interesting test to check their method. They obtained two samples of wood from Egyptian tombs: one from the tomb of Sneferu at Meydum, the other from the tomb of Zoser at Sakkara. Archaeological evidence indicated that both samples were about 4,600 years old. The carbon 14 measurement gave an age that agreed with this within the statistical counting error—a remarkable demonstration and a well-executed piece of scientific work.

Still another possible clock is the isotope potassium 40. This isotope may decay to calcium 40 or to argon 40. Too little is known about its half-life, however, to make this clock usable at the present time.

ULTIMATELY the most interesting question to be answered by our radioactive clock is: How old is the earth itself? Older estimates have varied from the 17th-century calculation of the Irish Archbishop James Ussher that the earth was formed at nine o'clock on the morning of October 12, 4004 B.C., to the almost equally specific figure of 1,972,949,048 years arrived at by the ancient Hindus. The Hindus, oddly, were in the right order of magnitude.

The oldest rocks that have been reliably dated are about two billion years old. But obviously there is little hope of arriving at an accurate figure for the age of the earth by direct measurement of the age of any mineral, for it can never be ascertained from such evidence alone how close to the beginning of the earth's time that particular mineral was formed. As more minerals have been measured, older and older ones have been found, and it is known that there are rocks more ancient than those dated two billion years ago.

A more direct and at present the most reliable method of arriving at the age of the earth is the study of the ratio between lead 206 and 207, daughters of uranium 238 and 235, respectively, in lead ores occurring in various parts of the earth. As we have noted, because of the difference in their half-lives the ratio of U-235 to U-238 has changed during the earth's history. There was more U-235 in a gram of uranium in the past than there is today, and the ratio can be calculated for all times in the past. In the discussion of the thorium and uranium clocks, it was pointed out that the proportion of the various isotopes in common lead was almost constant. But it is not quite: actually "common" lead is made up of a "primeval" lead, inherited by the earth at its beginning, and very small additions of daughter lead from the thinly distributed uranium and thorium that exists through the earth's crust. These small additions vary with the age of the common lead. Using these facts

MILLION YEARS	GEOLOGICAL UNIT OF TIME			EVENTS	LEAD-RATIO CONTROL POINTS	MAGNETITE-HELIUM RATIOS	STRONTIUM-RUBIDIUM RATIOS
					(MILLION YEARS)		
0	CENOZOIC ERA	CENOZOIC PERIOD	PLIOCENE EPOCH	Man appears.			
			MIOCENE EPOCH	Mammals at peak. Grazing types spread.		17, 15, 15, 21	
			OLIGOCENE EPOCH	Mammals evolve rapidly. Great apes.		45	
			EOCENE EPOCH	Modern mammals appear.		46, 43	
			PALEOCENE EPOCH	Archaic mammals dominant.	58	53, 51, 75, 51, 50	
100	MESOZOIC ERA	CRETACEOUS PERIOD		Dinosaurs, pterodactyls, toothed birds reach peak, then disappear. Small mammals. Flowering plants and hardwood forests.			
		JURASSIC PERIOD		Dinosaurs and marine reptiles dominant.		117, 120, 100, 120, 102, 132, 132	110, 150, 100
		TRIASSIC PERIOD		Small dinosaurs. First mammals. Conifers and cycads dominate forests.		169, 158, 174, 165	
200	PALEOZOIC ERA	PERMIAN PERIOD		Continental uplift and orogeny.	215		
		PENNSYLVANIAN PERIOD		Reptiles and insects appear. Spore-bearing trees dominate forests.		200, 205, 200, 225, 215 245	
		MISSISSIPPIAN PERIOD		Climax of crinoids and bryozoans.	255	240, 245	
300		DEVONIAN PERIOD		First amphibians. Brachiopods reach climax. First forests.		240 340 365, 340	300, 280, 200, 240, 300, 540, 270, 450, 540
		SILURIAN PERIOD		Widespread coral reefs. First evidence of land life.	350	375	
400		ORDOVICIAN PERIOD		Invertebrates increase greatly. Trilobites reach peak differentiation.	440		
500		CAMBRIAN PERIOD		First abundant fossils. Marine life only. Trilobites and brachiopods dominant.			
	PRE-CAMBRIAN					550, 500, 550, 620	

HISTORY OF THE EARTH in comparatively recent geologic times has been classically worked out by the sequence of rock layers. This sequence is now supported by measuring the products of radioactive processes in the rocks. Three columns at the right give the results of such measurements.

and comparing the proportions of lead 206 and 207 with lead 204 in a number of samples of lead ore of various ages and locations, the British geologist Arthur Holmes has computed the probable age of the earth at 3.35 billion years. This value appears to be the most reliable yet obtained. As the half-lives of U-238 and U-235 are determined more precisely, and as more isotopic analyses of lead are made, the figure may be modified, but it will probably not change very much.

If THIS be taken as the age of the earth, how much older is the universe? Twenty years ago astronomers, led by the British cosmologist James Jeans, were in favor of the "long" time scale, reckoning the age of the universe as some thousands of billions of years. The discovery of the "red shift" in the light from distant galaxies, which led to the expanding universe theory, and work on the dissolution of galactic clusters, recently summarized by Bart Bok of Harvard University, have tended to shorten the time scale in the minds of most astronomers. In general it seems that the astronomers now are more conscious of the youthfulness of the universe than of its antiquity. Their estimates today are of the order of two to three billion years.

Geologists and radiochemists believe that through the study of the present abundances of unstable isotopes and estimates of their probable abundances in primordial times, they may be able to determine a limit for the age of the universe. If all the universe was created suddenly at a certain point in time, presumably all of the nuclear species now known, both natural and artificial, and many more besides, would have been formed in the process. This time of creation of atoms, if indeed such an event occurred at a single time, has been called "nucleogenesis." The present evidence seems to place nucleogenesis only a short time before the origin of the earth, three billion years ago.

The fact that we now have well-defined limits to the age of the earth, and even some suggestion of the age of the universe itself, brings us face to face with the hardly avoidable consideration that the material universe was born in a violent event at a time in the not too distant past. Thus our radioactive clocks, by introducing calculable time, however great, into the history of the earth and the universe, give a reality to events which was not present when they were considered to have happened vague "eons ago."

The Author

P. M. HURLEY, assistant professor of geology at the Massachusetts Institute of Technology, was born in Hong Kong, China, in 1912. He received his undergraduate education at the University of British Columbia and spent three years as a mining engineer and geologist working in the gold mines of the British Columbia Rockies. During the war he came to the U. S. to serve on the National Defense Research Committee and then went on to his appointment at M.I.T. His special field of research is the application of nuclear physics to such geological questions as the age of the earth.

Bibliography

RADIOACTIVITY, THE EARTH'S HEAT AND GEOLOGIC AGE MEASUREMENTS. Robley D. Evans. Geological Society of America, Special Paper 36, pages 267-277; January 31, 1942.

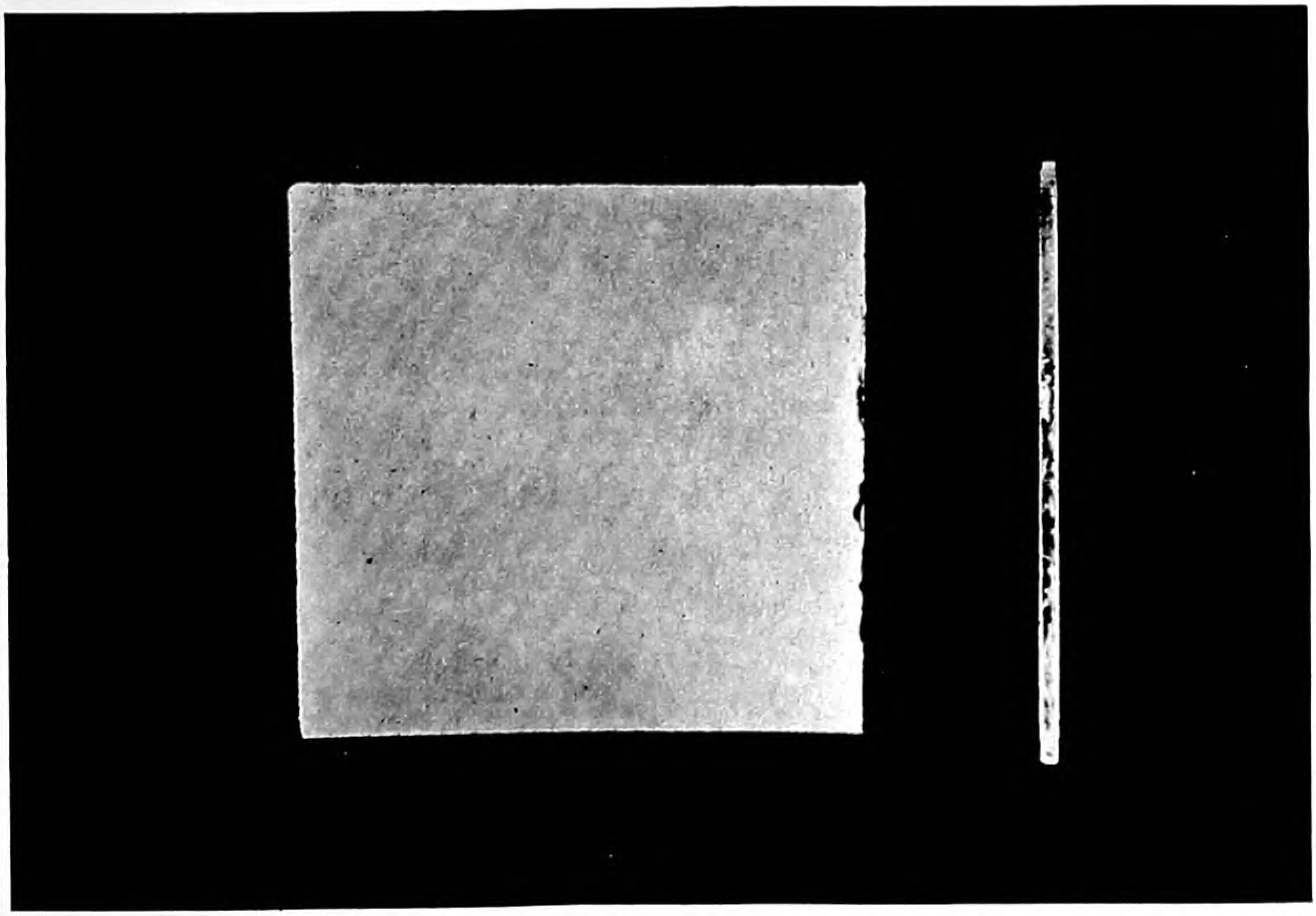
SCIENTIFIC AMERICAN August 1957, Vol. 197, No. 2, pp. 40-47 OFFPRINT 221

ELECTROLUMINESCENCE

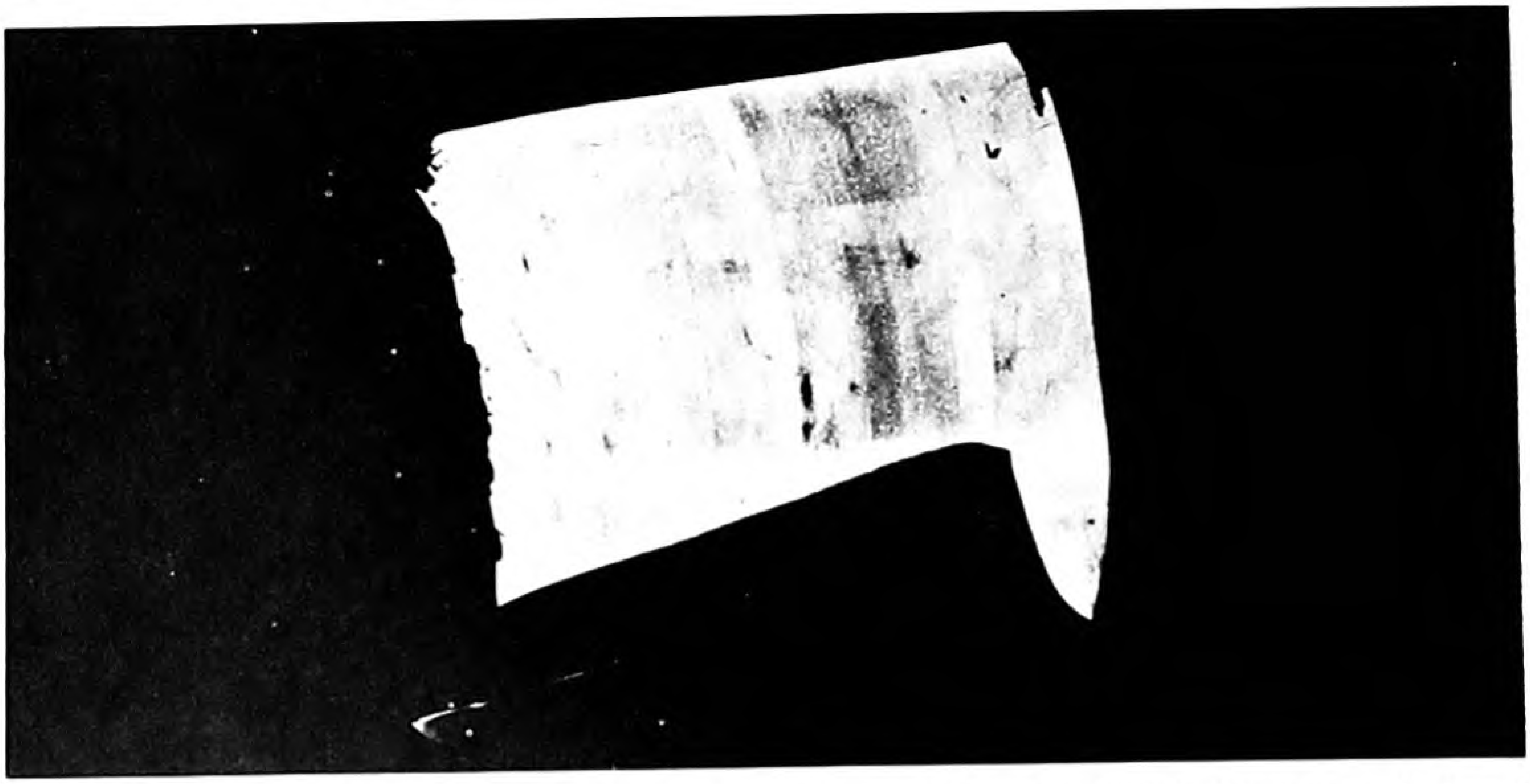
by Henry F. Ivey

When a luminescent substance such as zinc sulfide is placed in an alternating electric field, it emits light. This effect makes possible thin illuminating panels and many other useful devices.

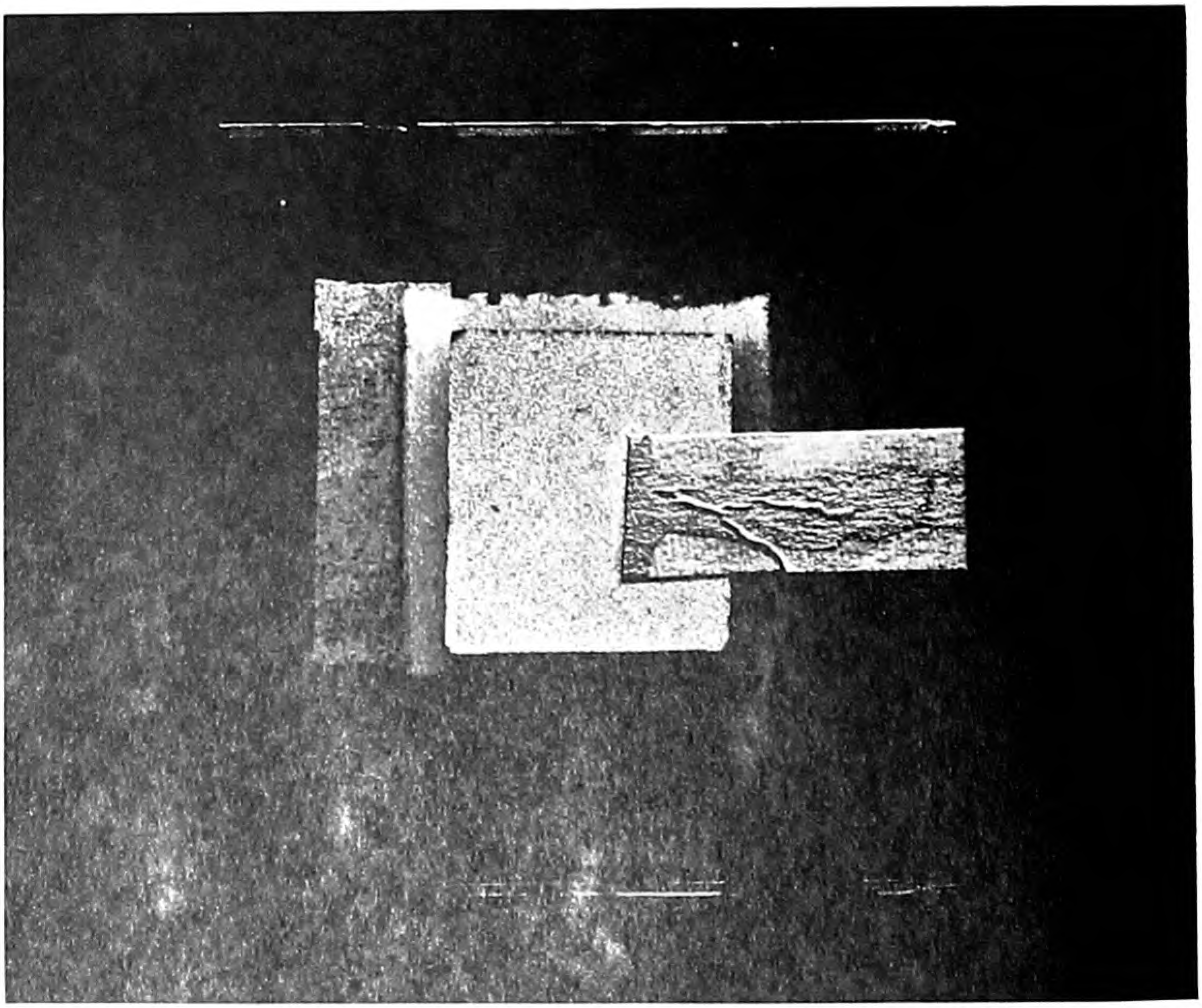
A new revolution in artificial lighting is in the making. Through most of human history man illuminated the darkness by means of a flame—progressing from the torch to the candle to the kerosene lamp. Within the last three quarters of a century our mode of living has been transformed by the incandescent and the fluorescent lamp. The innovation now under development is the luminescent panel. The prospect is that within a few years our houses and offices will begin to be lighted by luminous ceilings and walls. All sorts of attractive possibilities are opened up by this method of illumination. It will give us light of almost any color or brightness at will. Merely by turning a knob, we shall be able to bathe our rooms with rosy light on days of dreary skies, or with cool blue light when the sun is



EXPERIMENTAL ILLUMINATING PANEL is seen from the front and side. The phosphor of the panel is embedded in a thin layer of nonconducting material. In back of this layer is a thin layer of opaque conducting material. In front of it is a film of conducting material on glass. The alternating voltage which causes the phosphor to glow is imposed across the conducting layers.



FLEXIBLE PANEL in this photograph was made by spraying a layer of phosphor suspended in a nonconducting material on a screen with 200 meshes to the inch and covering it with a layer of vaporized aluminum. Panel operates at 110 volts and 4,000 cycles.



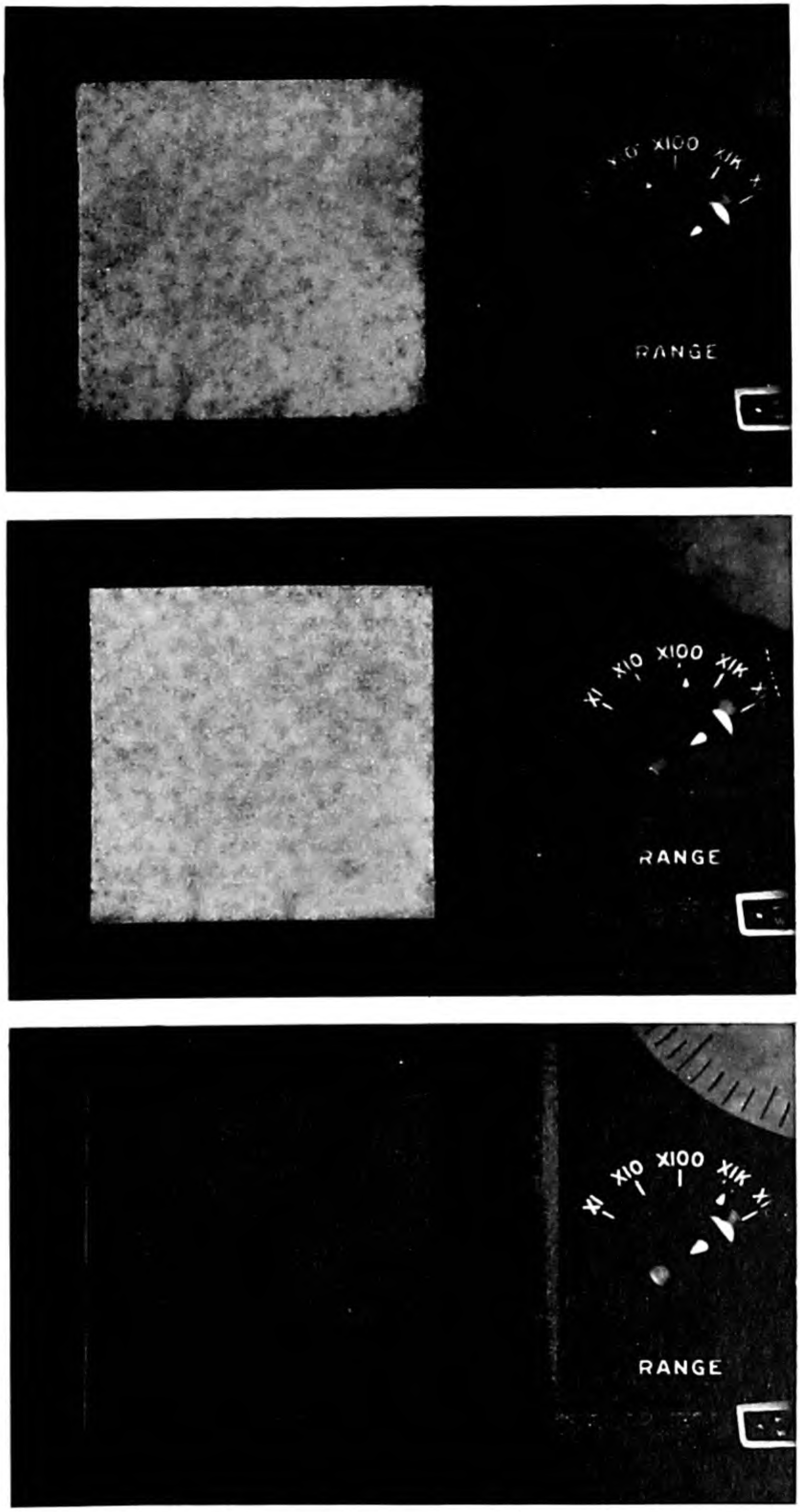
TEST PANEL is built up on glass. The oblong at right is the connection of the square electrode beneath it. The larger square beneath the electrode is phosphor in plastic. The oblong at left is the connection of the other electrode, which is not visible.

harsh. Night will be turned into day by curtains of light at our windows—draperies glowing with the golden shimmer of sunlight.

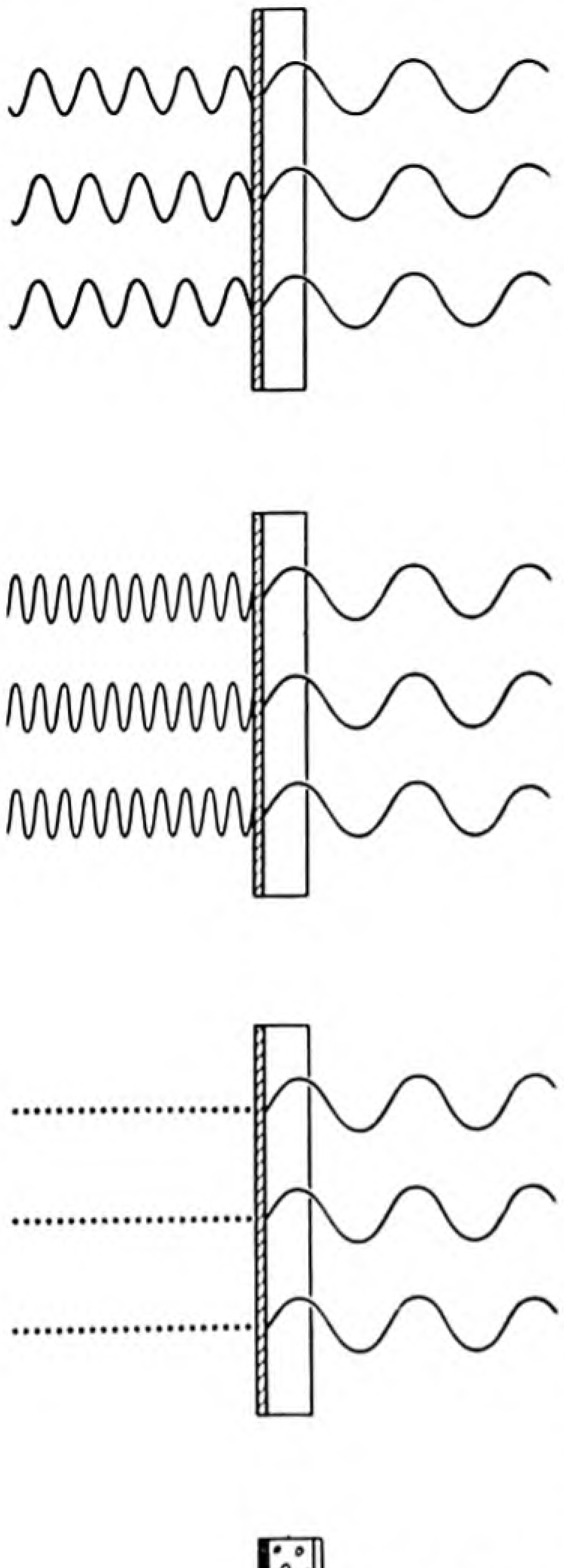
These, however, are only the more glamorous aspects of the phenomenon which is the subject of this article. Electroluminescence promises to revolutionize not only illumination but also techniques of picture-making. It can transform invisible radiations—ultraviolet, X-rays, cathode rays—into visible light. It may therefore improve X-ray photography and television, and introduce entirely new ways of forming pictures. But fully as interesting as its technological possibilities is the phenomenon itself. Electroluminescence is a new method of generating light. As such it gives us a new outlook on the behavior of energy and matter, in a realm which has recently become a cornucopia of useful and fundamental information—solid-state physics.

In a broad sense electroluminescence is not basically different from the luminescence we see in a fluorescent lamp, a fluoroscope or a television (cathode-ray) picture tube. In all of these devices the light is produced by exciting a phosphor. The difference among them lies in the agent used for excitation. In a fluorescent lamp the phosphor is excited by ultraviolet rays, produced in the tube by a gas discharge; in a fluoroscopic screen it is excited by X-rays; in a cathode-ray tube, by a beam of high-speed electrons. In the case of electroluminescence, the exciting agent is an alternating electric field. The phosphor, embedded in a thin layer of nonconducting material, is sandwiched between two layers of a material that conducts electricity. When an alternating voltage is applied to these outer layers (electrodes), it sets up an alternating electric field across the phosphor layer and causes the phosphor to emit light. From a practical point of view, the most important feature of this system is that it frees us from the confines of a tube or bulb in producing light. We can make light-emitting surfaces of any size or shape. Moreover, our "two-dimensional lamp" can be almost as thin as we please. The phosphor layer is only a few thousandths of an inch thick, and the outer layers of the sandwich can be fine wire meshes or films of a conducting substance (*e.g.*, tin oxide).

The fact that light could be generated by this method was discovered in 1936 by a Frenchman named George Destriau, and it is sometimes called the



COLOR OF PANEL CHANGES with frequency. Here the luminescent layer consists of two phosphors, one green and one red. The green phosphor is brighter at low frequencies; the red, at high frequencies. At intermediate frequencies the colors mix to produce white.



FOUR PHOSPHOR DEVICES are depicted schematically. In the fluorescent lamp (*top*) the phosphor (*hatched layer*) converts ultraviolet radiation (*wavy lines at left*) into light (*wavy lines at right*). In the fluoroscope (*second from top*) it converts X-rays into light. In the cathode-ray tube (*third*) it converts high-speed electrons (*dots*) into light. In the electroluminescent illuminating panel (*bottom*), the phosphor (*particles embedded in middle layer*) converts an alternating current (*wave symbol*) into light.

Destriau effect. He placed some zinc sulfide phosphors in an intense alternating field and made them glow—but so feebly that the light could be seen only in a dark room by eyes adapted to the darkness. The idea that phosphors could be excited by an electric field met with considerable doubt at first, but since 1950 Destriau's discovery has been followed up by intensive research, not only because it promises a new form of lighting but also because it is yielding new information about phosphors and the solid state.

The key to luminescence is the propensity of electrons in atoms to jump from one energy state to another. Upon absorbing a quantum of energy from some outside source, an electron jumps to a higher, or "excited," state. It may then jump back to its original state, re-emitting the absorbed energy. In a phosphor, part of this energy is emitted as light. But the process is complex, and certain conditions must be met to make the electron jumps produce light rather than merely heat or other radiations.

Here is a phosphor crystal, say zinc sulfide. Since its closely packed atoms interact with one another, we must think of its electrons as occupying a band of energy levels, instead of the single, sharply defined energy level of an electron in an isolated atom. Nevertheless, just as a single electron in an atom is permitted only certain energy levels and is forbidden any intermediate energies, so there is a forbidden band of energy values which cannot be occupied by any electrons of the zinc or sulfur atoms. When the crystal is unexcited, the lowest energy band is completely filled with electrons. If exciting energy is supplied to the crystal, some of the electrons will jump from this band across the forbidden band just above it into a permissible higher band. Now if an electron jumps back to the lowest band, it will emit energy—but not in the form of light. The quantum of energy required for a jump from one level to another depends on the width of the gap, and the size of the quantum determines the frequency, or wavelength, of the radiation emitted: the bigger the quantum, the higher the frequency. In this case the gap across the forbidden band is so wide that an electron jumping back to the lower band emits its energy at a frequency in the ultraviolet region rather than in the lower-frequency region of visible light.

But we can narrow the gap. Suppose we insert a new rung within the forbidden zone to shorten the step for the jumping electrons. We can accomplish this by introducing a few copper atoms into the zinc sulfide crystal. An energy level permissible to copper's electrons lies within the band forbidden to those of zinc and sulfur. If a copper electron is excited to a higher energy level, it leaves a vacancy which can be filled by an electron dropping from the higher band. The drop to this level is short enough—that is, the quantum is of the right size—to produce emission of light. The electron berths provided by the copper atoms are therefore called luminescence centers.

There are other holes, within the forbidden band and close to the higher energy band, into which electrons may fall. They are called traps, and are created by imperfections in the crystal or by chlorine atoms, which are introduced to control the number of copper atoms dissolved in the crystal. Electrons falling into traps generate only heat, but they are easily excited out of the traps and then can fall into empty luminescence centers, producing light.

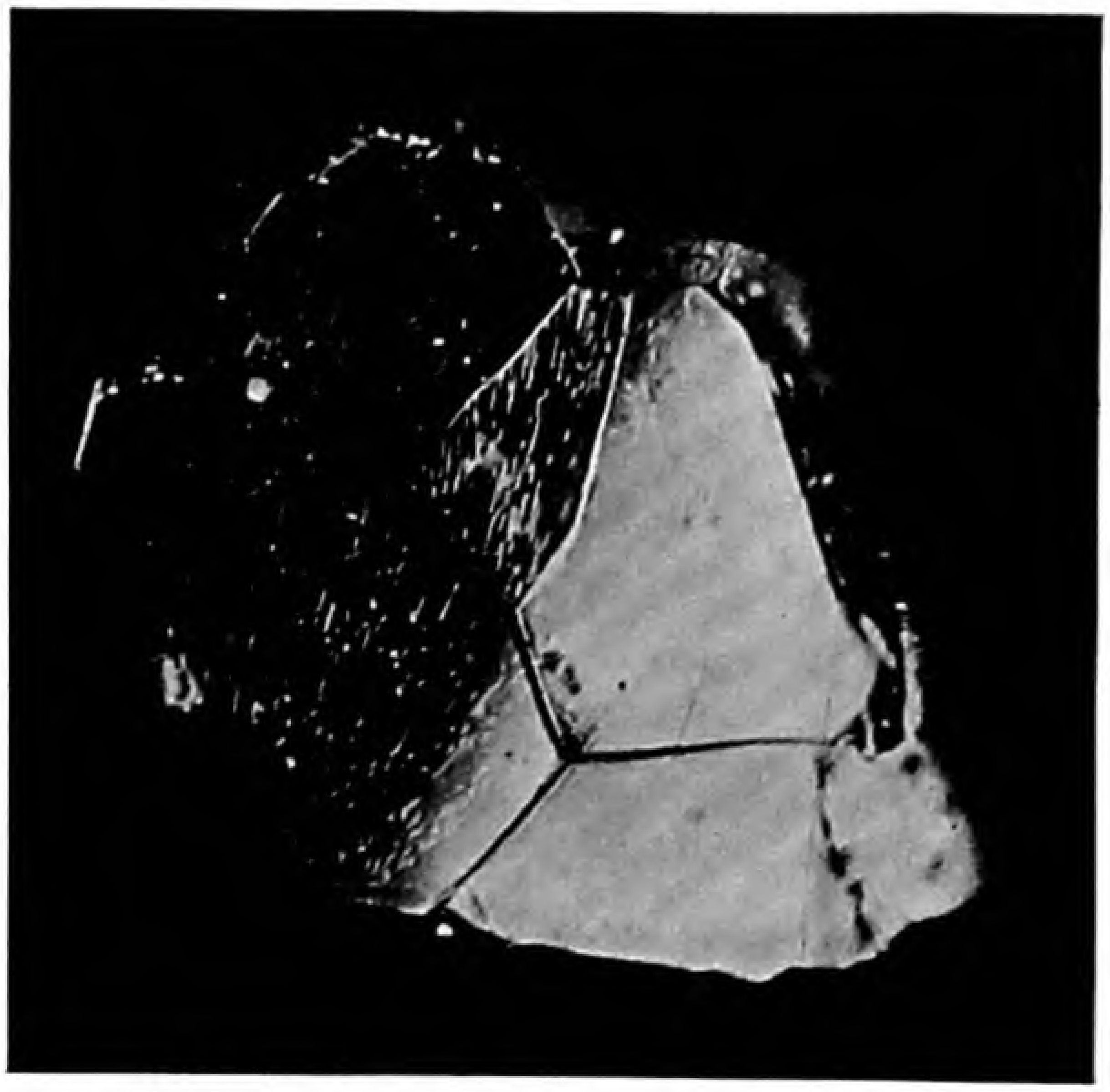
The foregoing describes the mechanism of the more familiar types of luminescence, such as fluorescence. The ultraviolet rays in a fluorescent tube excite electrons contained in the phosphor to the higher permitted energy band, and the excited electrons then fall into empty luminescence centers. When we come to electroluminescence, the process is more indirect. The alternating electric field cannot kick electrons out of the centers directly. What it does is to accelerate free electrons, and these accelerated particles then knock electrons out of the luminescence centers, creating vacancies into which electrons can fall with emission of light. In practice, if the method is to produce much light, the phosphor has to be specially treated to increase the effectiveness of the electric field. Small particles of copper sulfide are introduced into the phosphor. Because copper sulfide is a good conductor, each of these particles concentrates and intensifies the electric field around it, and so enhances the acceleration of electrons.

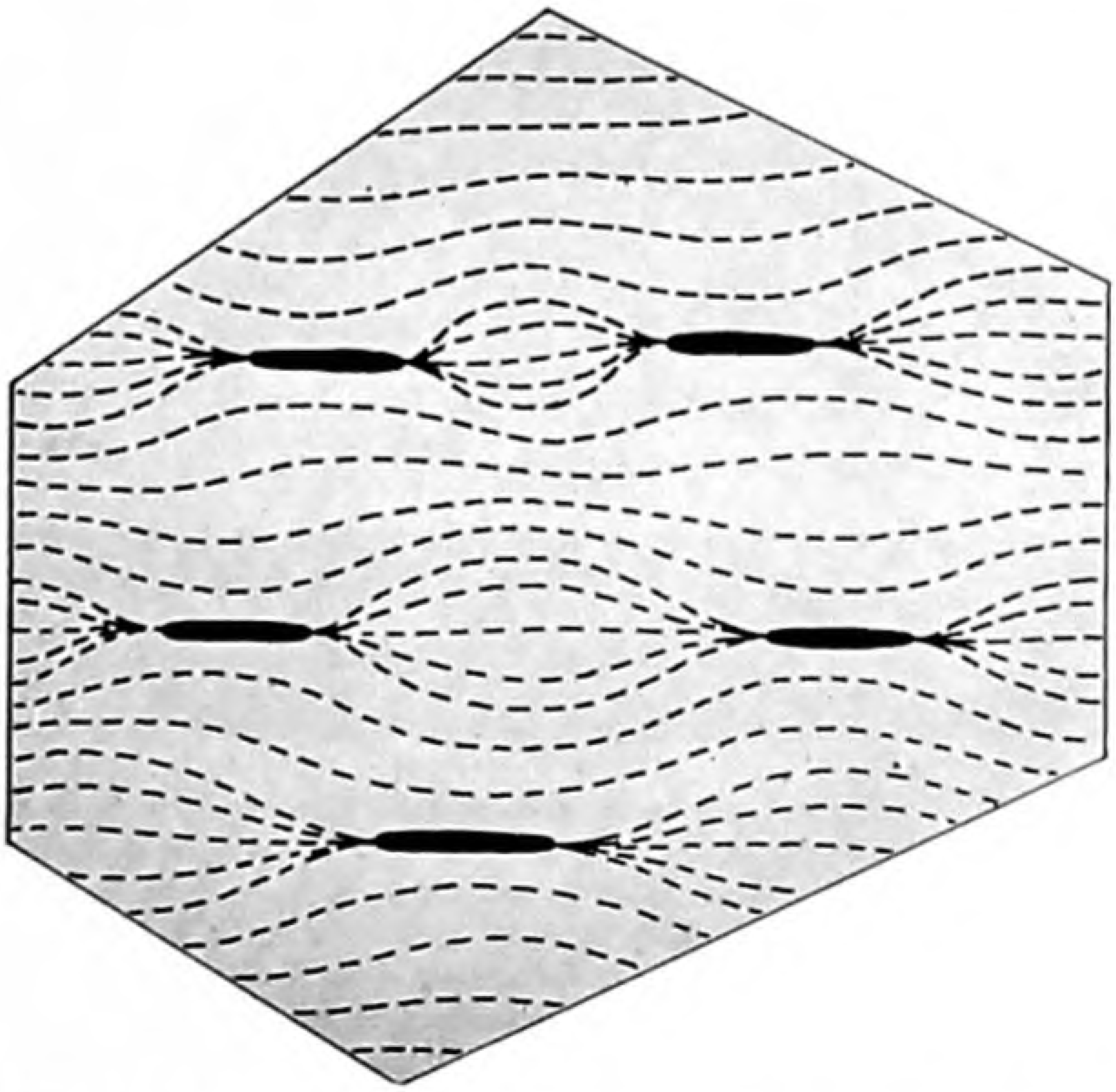
We get something of a picture of how the process works by setting up a mechanical analogy. Let us take a board pitted with holes like a Chinese checkerboard: deep holes represent luminescence centers and much shallower ones represent traps. Using marbles as electrons, we fill all the luminescence centers and some of the traps with electrons. The board is mounted on a fulcrum so that it

EXPERIMENTAL ROOM at the Bloomfield, N. J., plant of the Westinghouse Electric Corporation is illuminated with electroluminescent panels. The panels were made with zinc sulfide activated with copper. They operate at 400 volts and 3,000 cycles.

tilts from one side to the other like a seesaw: this represents the alternating electric field [*see drawings on page 166*].

Now suppose we look at one side of the board as it is tilted down. As the steepness of the tilt increases, marbles begin to roll out of the shallow traps; some hit marbles in luminescence centers and knock them out; soon there is an avalanche of marbles rolling down the surface of the board (representing the higher energy band). Then, after the board reaches its steepest tilt and starts back up again, the avalanche slows down. When the board returns to the horizontal position, the marbles stop rolling. They fall into traps, because there are no empty luminescence centers out near the periphery where they are concentrated. Many of the centers nearer the middle of the board are, however,





CRYSTALS OF CADMIUM SULFIDE, enlarged 200 times in the photograph at left, are shot through with needle-like dislocations of their crystal lattices. The diagram at right shows a crystal of zinc sulfide containing particles of copper sulfide in similar dislocations. Because copper sulfide is a good conductor, the particles distort and intensify an electric field (*broken lines*) around them.

unoccupied. Now the side of the board that we are watching begins to tilt upward. The marbles roll out of the traps and back toward the empty centers. They drop into these holes. The cycle is completed when the board again comes back to the horizontal position.

Meanwhile on the other half of the board the same thing has happened, but one half-cycle out of phase. Thus while luminescence centers are being emptied on one side of the board they are being filled on the other.

This is what happens around each local intensification of the electric field in a phosphor. The field, though not strong enough to dislodge electrons from luminescence centers, can free electrons from the traps and so start an avalanche. Once started, the avalanche excites many of the electrons out of the centers by means of collisions. It slows down as it reaches areas of low field strength. The electrons then fill nearby traps, only to be liberated again by the reverse build-up of the alternating field, which accelerates them back toward the empty centers. When electrons drop into the centers, they emit light.

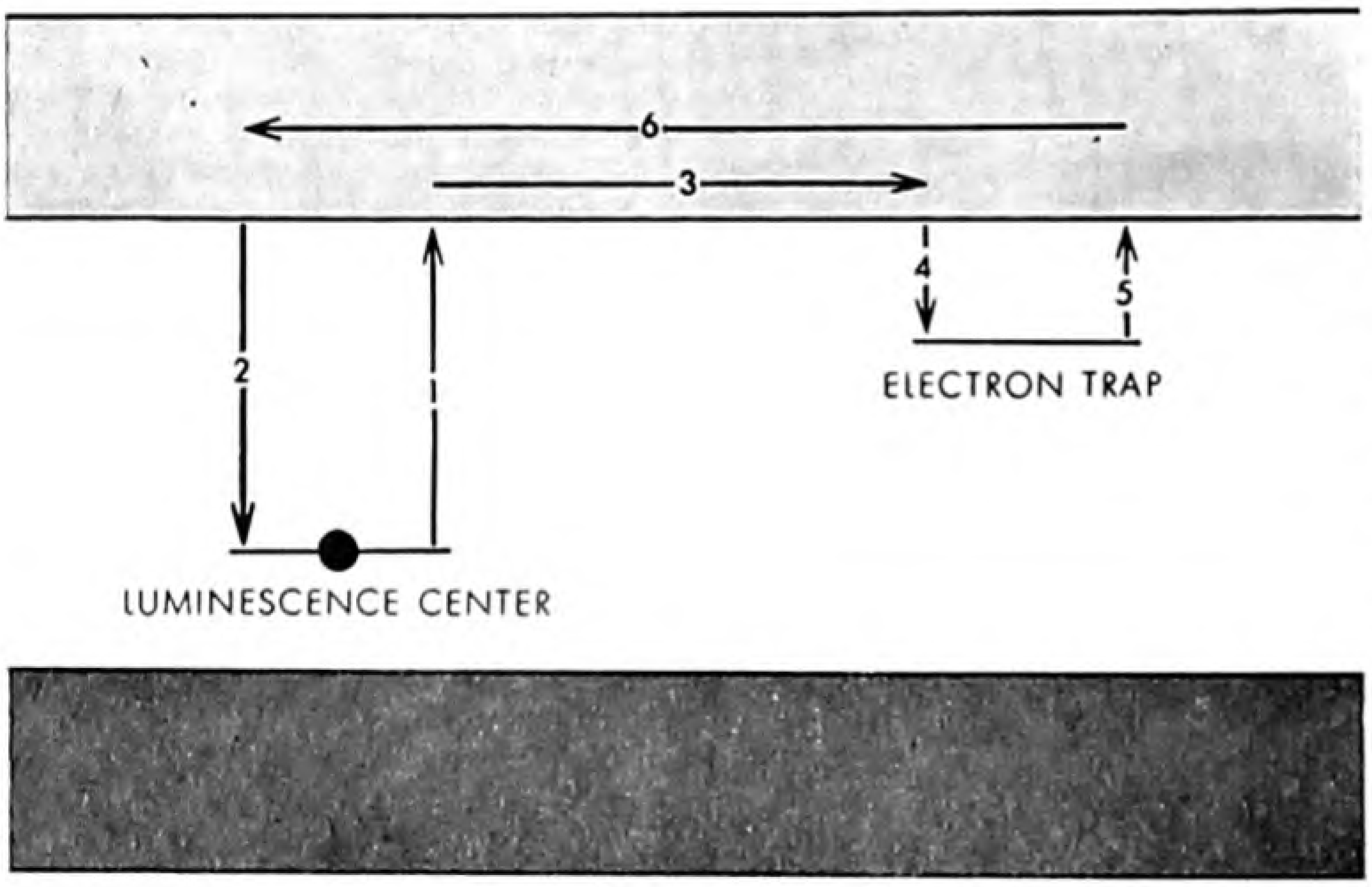
The main problem in electroluminescence is to get enough output of light from the phosphor. There are two ways to step up its brightness: increase the voltage (*i.e.*, strengthen the electric field) or speed up the field's oscillating frequency. If we strengthen the field, we generate more light emission by emptying and refilling more luminescence centers in each cycle; if we increase the frequency (rock the board faster), we get the same result by emptying and refilling the centers more rapidly. But there are limits to both stratagems. Too strong a field will break down the insulating property of the phosphor layer; too short a cycle will not give the electrons time enough to emerge from their traps.

The highest brightness yet produced—achieved by exciting a green phosphor with a field at 20,000 cycles per second—is 2,500 lumens per square foot of phosphor area. This is brighter than a standard 40-watt fluorescent lamp. But it takes an inordinate amount of electric power, because the efficiency of the electroluminescent process is low. At 20,000 cycles per second electroluminescence yields only about one lumen per watt of applied power, whereas a fluorescent lamp gives 65 lumens per watt, and even an incandescent electric-light bulb (of 100 watts) yields 16 lumens per watt.

The reason for the low efficiency of electroluminescence is that the avalanching electrons spend most of their energy in collisions with atoms in the crystal: only a minority knock electrons out of luminescence centers. It has been found that the process behaves most efficiently when the applied electricity has a potential of a few hundred volts and a frequency of a few hundred cycles per second. The highest efficiency obtained so far is 12 lumens per watt. The aim of the workers on this problem is to raise the efficiency to about that of diffused fluorescent lighting—namely, about 30 lumens per watt.

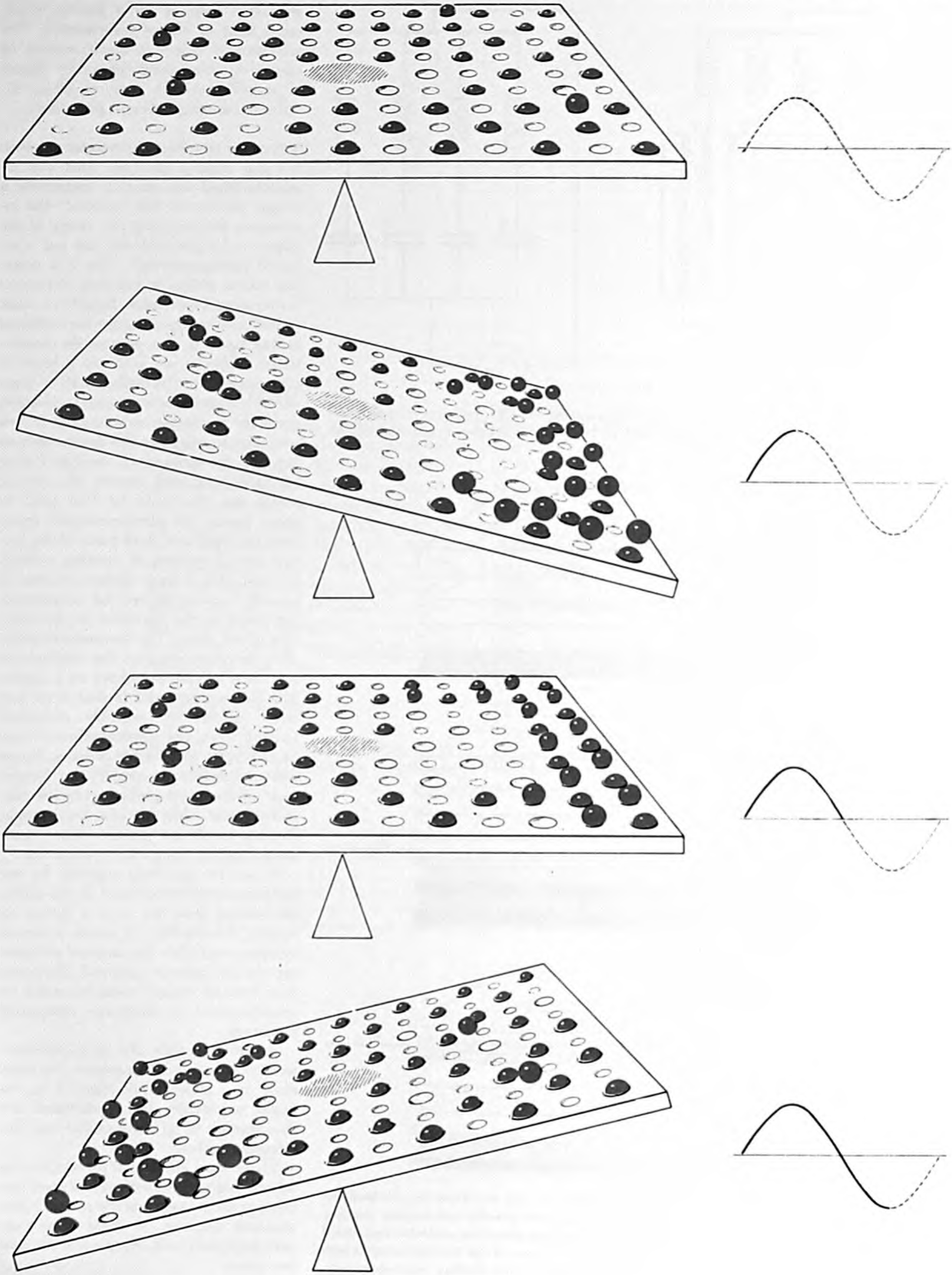
Until more effective materials are developed, electroluminescent lighting will require alternating electricity of considerably higher frequency than the 60-cycle current now supplied by power lines. For houses, small converters using transistors may serve to convert this current to the necessary frequencies. In a large office building it would be quite feasible to install a generator which would deliver power at frequencies around 1,000 cycles per second. The cost of such an installation would be offset by a great saving in the cubage of the building, because it would need only very thin ceilings, in place of the space now required for lamp fixtures. Whereas recessed fluorescent lamps, for instance, need a depth of 14 inches, electroluminescence would thin the ceiling down to less than one inch.

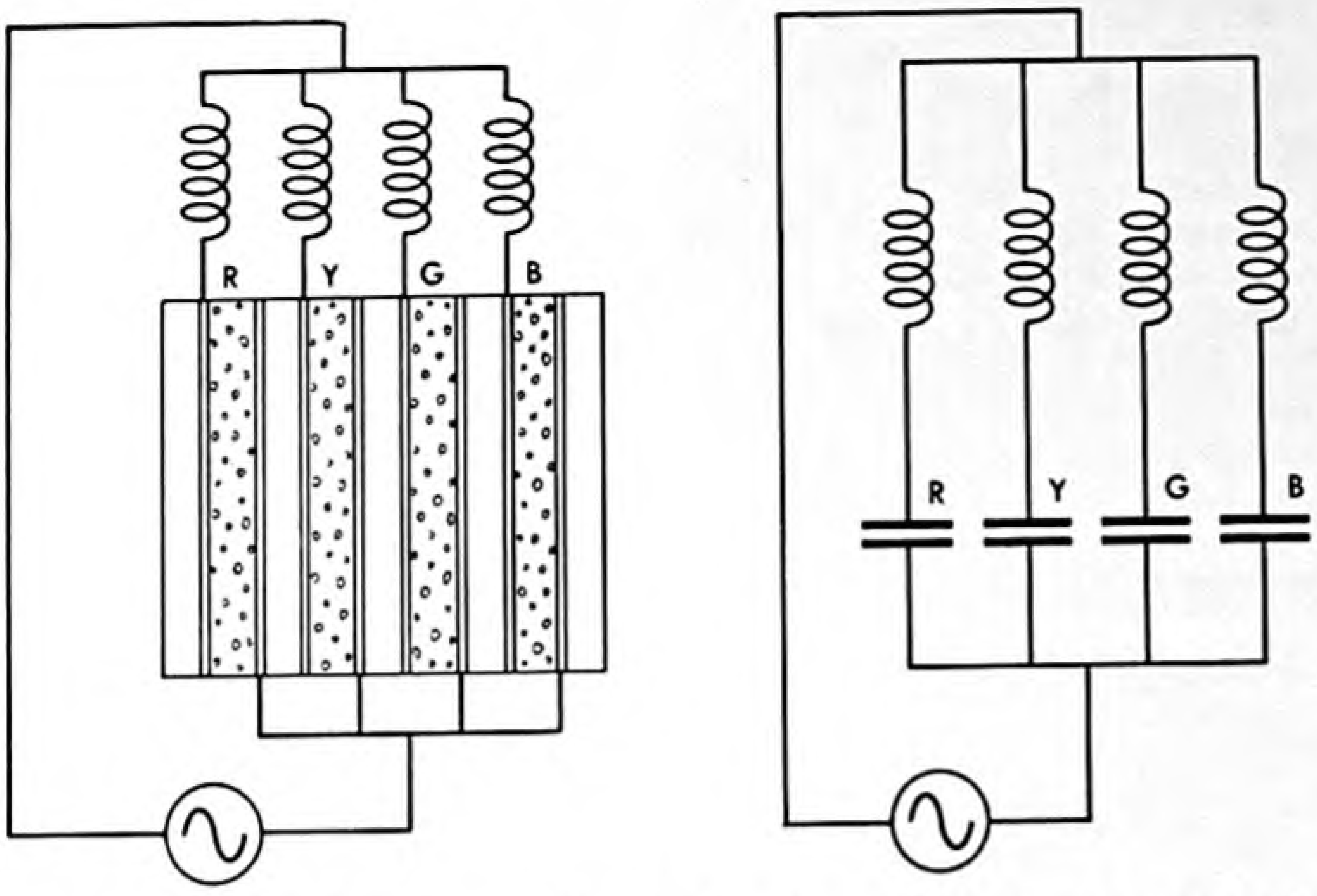
The fact that electroluminescent surfaces can take any size or shape provides an endless list of possibilities. Light might be built into walls, doors, stairs, balustrades, domes or any other elements of a building. It could be built into furniture and even into draperies, because the luminescent sheets can be made as flexible as cloth. And the color of the light could be changed instantly by tuning the applied electricity to a different frequency. The available colors depend on the substances introduced into the phosphor to provide luminescence centers. Copper gives blue or green. Manganese, whose electrons occupy an energy level closer to the upper energy band of the phosphor (*i.e.*, are



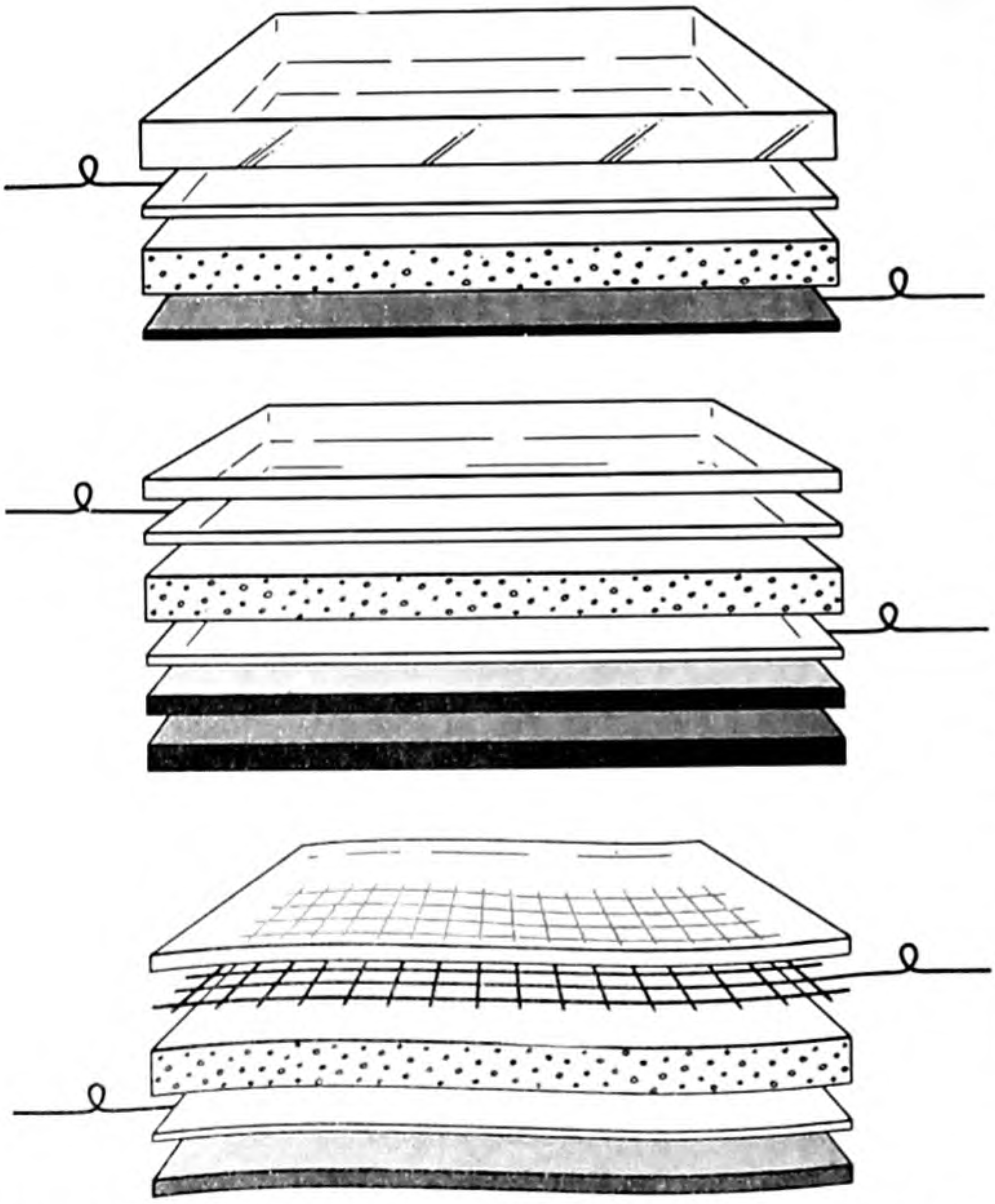
ENERGY BANDS of a zinc sulfide crystal are schematically depicted. At bottom is the band occupied by unexcited electrons; at top, the band occupied by excited electrons. Between the two is a "forbidden" band which cannot be occupied by electrons in a pure crystal. If an electron in such a crystal jumps from the lower band to the upper and falls back again, it may emit ultraviolet radiation or simply radiate heat. If copper atoms are introduced into the crystal, they form a luminescence center whose energy level is in the forbidden band. If an electron jumps from this level (1) and falls back again (2), it emits light. The electron can also wander through the crystal (3) and fall into an electron trap (4). Such a trap is created by an imperfection in the crystal or by chlorine atoms which are introduced to control the number of copper atoms. The electrons are easily dislodged from the trap (5) and can then fall back to the level of the luminescence center (6 *and* 2).

MECHANICAL ANALOGY discussed in the text of this article is depicted on the opposite page. The colored balls are electrons; the deep holes, luminescence centers; the shallow holes, electron traps. The hatched area in the center represents the copper sulfide introduced into the zinc sulfide crystal to intensify the electric field. The tilting of the board is an analogy for the alternation of the voltage, which is also represented by the curves at the far right.





MULTICOLORED PANEL of one type is made by stacking red (R), yellow (Y), green (G) and blue (B) phosphors. Inductors are used to tune in each layer (*circuit at right*).



COMPLEX PANELS are dissected. Layers of the panel at top are, from top to bottom: glass, conducting coating, phosphor particles suspended in plastic, conducting coating. Layers of the cell in middle are: plastic, conducting coating, phosphor particles embedded in low-melting-point glass, reflecting enamel, metal base. Layers of the flexible panel at bottom are: plastic, wire mesh, phosphor embedded in plastic, conducting coating, plastic.

separated from it by a shorter gap), emits light of longer wavelengths. The selection of colors is made subject to control by frequency tuning by means of several methods, some of which involve combining different phosphors.

Let us see now how electroluminescence can form a picture. This can be accomplished by what is essentially a simple device. As the "camera"—the instrument for receiving the image of the object to be pictured—we can use a so-called photoconductor. This is a material whose ability to conduct electricity is increased by exposure to light or other radiation—the more intense the radiation falling upon it, the greater its conductivity. Suppose we back up a layer of photoconducting material with a layer of electroluminescent material and apply a voltage across the combination. If now we focus an image on the photoconducting side, the intensity of the light falling on each point will govern the voltage across the two layers at that spot. In other words, the photoconductor translates the light and dark parts of the picture into a pattern of varying voltage. In turn, this voltage pattern excites in equally varying degree the corresponding points in the phosphor on the other side of the sheet. The luminescent phosphor therefore displays the original image, as if it were projected on a screen. The interesting point is that even patterns formed by invisible radiations (acting upon the photoconductor) can be converted into visible pictures. Moreover, the device can amplify the "brightness" of the image picked up by the photoconductor. Thus it could, for example, display an X-ray picture many times more brightly than a fluoroscope can.

If we let the light emitted by the luminescent side feed back to the photoconducting side, we have a device for storing information: it stores a certain voltage even after the original stimulating signal has been removed. This compact form of storage could be useful for incorporation in electronic computing machines.

It seems possible that an electroluminescent screen could perform the functions of a cathode-ray tube; if so, we might eventually have television sets thin enough to be hung on the wall like a framed picture.

There are a number of problems to be solved before electroluminescence enters our daily lives, but the research laboratories working on them hope that, with luck, they will be solved in the next few years.

The Author

HENRY F. IVEY is in charge of the phosphor section of the Westinghouse Electric Corporation. He was trained in physics at the University of Georgia and the Massachusetts Institute of Technology, where he earned his Ph.D. in 1944. As a staff member of the M.I.T. Radiation Laboratory during World War II, Ivey worked on cathode-ray screens for radar. This led him to research on the luminescence of television screens and on the new lighting technique which he describes in this article.

Bibliography

THE ELECTRICAL PROPERTIES OF PHOSPHORS. G. F. J. Garlick in *Progress in Semiconductors,* Vol. 1, pages 101-133; 1956.

ELECTROLUMINESCENCE AND RELATED TOPICS. G. Destriau and H. F. Ivey in *Proceedings of the I.R.E.,* Vol. 43, No. 12, pages 1,911-1940; December, 1955.

THE NEW PHENOMENON OF ELECTROPHOTOLUMINESCENCE AND ITS POSSIBILITIES FOR THE INVESTIGATION OF CRYSTAL LATTICE. G. Destriau in *The Philosophical Magazine,* Ser. 7, Vol. 38, No. 285, pages 700-739, October, 1947; Vol. 38, No. 286, pages 774-793, November, 1947; Vol. 38, No. 287, pages 880-888, December, 1947.

SOLID-STATE LUMINESCENCE. Ferd E. Williams in *Advances in Electronics,* Vol. 5 pages 137-168; 1953.

SCIENTIFIC AMERICAN November 1958, Vol. 199, No. 5, pp. 31-37 OFFPRINT 222

THE REVIVAL OF THERMOELECTRICITY

by Abram F. Joffe

The Seebeck and Peltier effects dispense with moving parts in the transformation of energy from heat to electricity. Long regarded as curiosities, they are now generating a technical revolution.

Farms and pioneer settlements in many regions of the U.S.S.R. remote from electrical power grids today enjoy the advantages of electric light, electrified machinery and radio, thanks to a new kind of generator. This generator has no moving parts, but converts the heat of a kerosene lamp or a wood stove directly into electrical power. In the cities of the nation, apartment dwellers are soon to be furnished with an inexpensive refrigerator which operates on a quite opposite effect. This refrigerator has no motor or compressor and no refrigerant fluid; it uses electrical energy directly to pump the heat out of its interior. A variation of this effect applied to the heating of indoor working and living spaces promises to achieve great economy in the use of electrical energy for such purposes.

In the modern world, developments of this scope are never the exclusive province of one country. When I visited the U. S. earlier this year, I found scientists and engineers fired with these possibilities and working actively to realize them. Evidently we stand on the threshold of a new era in power engineering, heating and refrigeration. This prospect flows from the now-rapid advance of thermoelectricity: the direct transformation of thermal energy into electrical energy and the reciprocal transformation of electrical energy into heat.

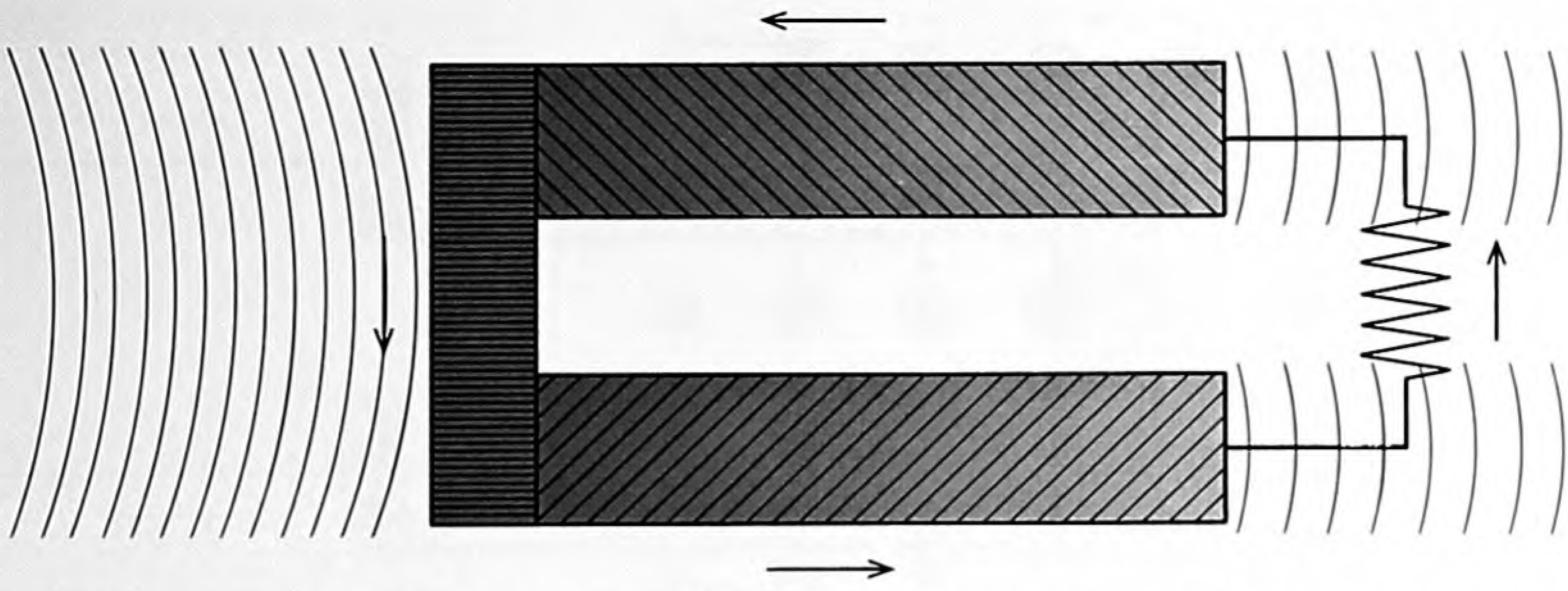
Thermoelectricity is not a new science. It is, in fact, just as old as our knowledge of the electromagnetic effects upon which electrical technology is based. The history of electromagnetism began in 1820, when Hans Christian Oersted reported his observation that a magnetic needle is deflected by the flow of an electric current in a nearby conductor. Only a year later Thomas Johann Seebeck reported his observation that a magnetic needle held near a circuit made up of two different conductors is deflected when part of the circuit is heated. Unhappily this promising discovery was obscured by the discoverer's own misjudgment. A comparison of the Oersted and Seebeck observations suggests naturally that a current was flowing in Seebeck's circuit. That is the way Seebeck's contemporaries interpreted his discovery. But, alas, Seebeck did not. He thought he had shown that magnetization may be caused by a difference in temperature, and attempted to explain terrestrial magnetism by the temperature difference between the Equator and the poles.

Seebeck's error had two different kinds of consequences. In trying to convince his opponents that his effect was magnetic rather than electric in origin, and thus not related to the electrical properties of the substance, Seebeck investigated an enormous number of different materials, not only metals but also metal oxides, minerals and other compounds. Among these were substances which we now call semiconductors and which we use today in thermoelectric generators.

On the other hand, in his unyielding battle against electric current, Seebeck succeeded in discouraging himself and others from attempting to use his thermoelectric effect as a source of current. Yet it has been found that the substances he used in his experiments can transform heat into electricity with an efficiency of about 3 per cent. This compares favorably with the mechanical efficiency of the steam engines of his day. In the laboratories of the time, moreover, the only machines for producing electric current were extremely weak electrostatic generators. Not until half a century later was electricity successfully generated by electromagnetic machines driven by steam engines.

It is hard to say how the history of electrical engineering and electronics might have unfolded had Seebeck's discovery been wisely used. For a long time it would have provided the best source of electric energy. But this is not what happened. The rapid development of electromagnetism diverted the interest of succeeding generations of physicists away from thermoelectricity. Certainly one could expect no significant thermoelectric effects from metals, and these were the only conductors used in electrical engineering for the next 100 years. Seebeck's mineral semiconductors were ignored even in the design of thermocouples for temperature measurement, the sole application of thermoelectricity throughout this period. The Seebeck effect became a curiosity, relegated to the last pages of physics textbooks.

For a century the marvelous possibil-



SEEBECK EFFECT brings about conversion of heat into electric energy: heat input to "hot" junction (*left*) of two dissimilar semiconductors causes current to flow from hot to cold end of semiconductor (*n*-type) at bottom and from cold end to hot in semiconductor (*p*-type) at top. An electric current is thus made to flow through the entire circuit, including the output element at right.

ities of thermoelectricity remained asleep, like the princess in the fairy tale.

This slumber was momentarily disturbed in 1834. A French watchmaker, Jean Charles Athanase Peltier, discovered that the passage of a current through the junction between two different conductors is associated with a thermal effect. But like Seebeck, Peltier failed to understand his discovery. He saw it only as showing that Ohm's law may be disobeyed by sufficiently weak currents. The true nature of Peltier's discovery was clearly demonstrated in 1838 by Emil Lenz, a member of the St. Petersburg Academy. He placed a drop of water on the junction; when he made the current flow in one direction, the water froze; when he made it flow in the other, the ice melted.

Could anyone fail to see that the junction liberated heat when the frozen drop melted and that it absorbed heat when the drop froze? Peltier's discovery ought to have found important applications, since it opened a new way to obtain cold and heat. But this discovery also spent 100 years in the kingdom of sleep.

The true meaning of Seebeck's and Peltier's discoveries was appreciated by scientists before the middle of the 19th century. Until the 1930s, however, attempts to apply them met with no success. This situation has changed radically in the last few years. What is it that has happened? What has ended the century-long sleep of thermoelectricity?

The prince who has awakened the princess is the semiconductor.

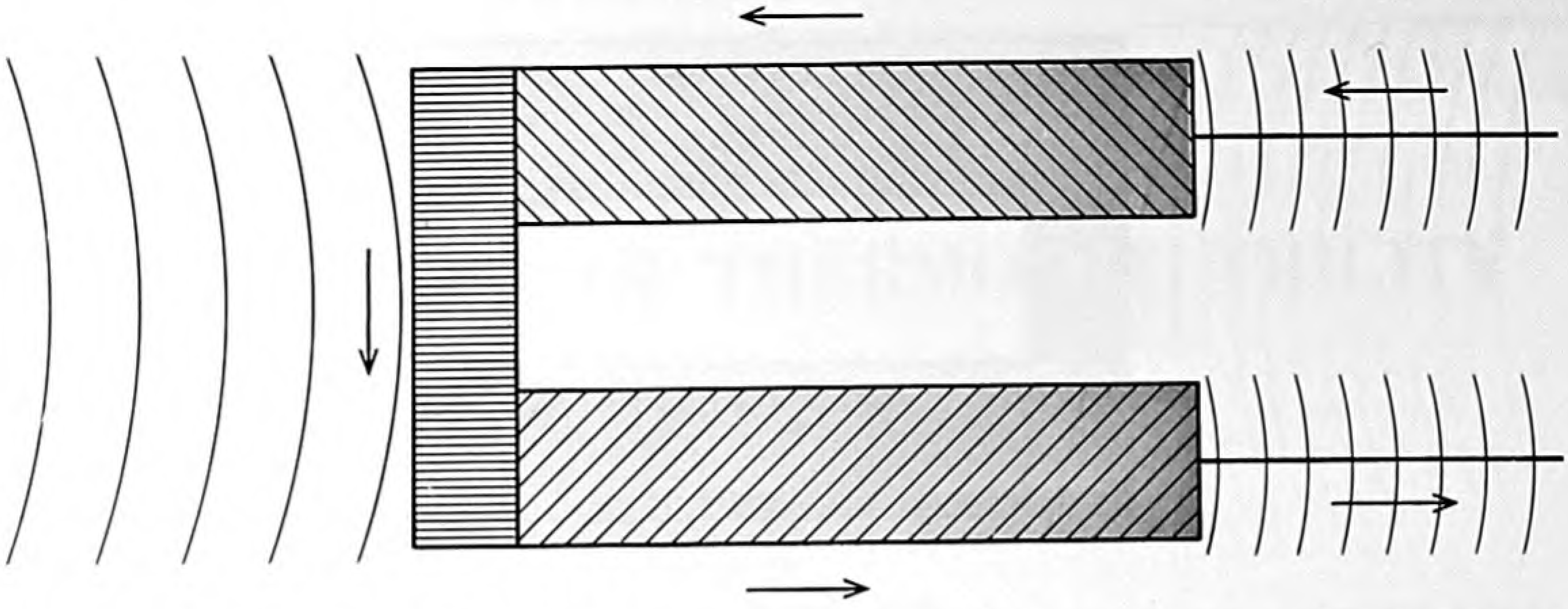
In 1926 a U. S. engineer named Lars O. Grondahl showed that an oxidized copper plate conducts an electrical current easily in one direction, but offers a very high resistance in the other. Thus if an alternating current is passed through such a plate, the current will flow for all practical purposes only in one direction. In effect the plate becomes a rectifier. Soon afterward it was found that, when such a plate is illuminated, a current is produced. The discovery of these properties in copper oxides attracted the attention of physicists to the large class of materials with low electrical conductivity which we now know as semiconductors. Investigation soon disclosed that such substances possess many other remarkable properties unknown in metals. Upon these properties are based the transistor, the solar battery and related developments in technology. In addition, it was found that thermoelectric effects are an order of magnitude larger in semiconductors than in metals.

This observation raises some fundamental physical questions that lie at the heart of future developments in thermoelectricity. Why is it that semiconductors have such a high thermoelectric advantage over metals, the "full" conductors of electric current? Why is it that semiconductors alone make thermoelectricity a fruitful realm of technology? Complete and exact answers to these questions require an excursion into quantum theory. One of the basic properties of this theory, however, is that it cannot be represented by a simple model. In attempting to make quantum theory easy to understand, we necessarily misrepresent it. For this reason I shall restrict this explanation to more familiar concepts that can be stated in the language of everyday experience.

We think of a current in a metal as the flow of electrons. In a metal each atom contributes at least one electron able to move freely within the metal. In semiconductors, on the other hand, only a very few atoms release such free electrons. The number of electrons available for current flow in a semiconductor is hundreds or thousands of times less than in a metal, and this accounts for the low conductivity of these substances.

When one end of a semiconductor—or a conductor, for that matter—is hotter than the other, electrons leave the hot end more often than they do the cold end. They tend to flow toward the cold end, and, since they are all negatively charged, the cold end soon becomes charged negatively with respect to the hot end. Now it is well known that like electric charges repel one another and unlike charges attract one another. The negatively charged cold end thus begins to repel the electrons flowing from the hot end. After a short time the flow of electrons from the hot end reaches equilibrium with the return flow from the cold end. The charges no longer accumulate, but the cold end remains negatively charged. The fewer the electrons available for the return flow, the higher will be the voltage attained at the cold end before equilibrium is reached. Since the number of free electrons is much smaller in a semiconductor, a temperature difference in a semiconductor will produce a much greater voltage than in a conductor. This is the essence of the thermoelectric advantage possessed by semiconductors.

Different semiconductors develop greater or lesser voltages. In all semiconductors the voltage increases with



PELTIER EFFECT, the reverse of the Seebeck effect (*opposite page*), makes thermoelectric cell act as refrigerating or heating unit. Electric input to the cell as shown here cools the junction at left, causing it to absorb heat, and heats the ends of the semiconductors at right, causing them to give up heat. By reversing the direction of the current the junction can be made to give up heat.

the difference in temperature between the hot and cold ends. The voltage across a given semiconductor when one of its ends is warmer than the other is the measure of its characteristic thermoelectric power, which is expressed in volts per degree centigrade. Semiconductors display thermoelectric power some hundreds of times greater than that of metals. But since metals develop only a few millionths of a volt, the thermoelectric power of semiconductors is still very small. Even when the difference in temperature of the two ends of a semiconductor is several hundred degrees, the semiconductor develops only 10ths of a volt. This, however, is enough to make thermoelectricity useful.

Semiconductors possess still another thermoelectric advantage not found in metals at all. In some types of semiconductor material the voltage differential between the hot and the cold end is set up not by the flow of negatively charged electrons but by the flow of positively charged "holes" vacated by electrons. As a result, the cold end in such a semiconductor becomes positively charged. The two types of semiconductor are designated as "*n*-type" (hot end positive) and "*p*-type" (cold end positive). In both types, of course, the direction of the current (electron flow) is from the positive to the negative end, as inside a battery [see "The Junction Transistor," by Morgan Sparks; SCIENTIFIC AMERICAN, July, 1952].

Let us now construct a thermoelectric circuit to generate an electric current. We take two semiconductors of opposite types, an *n*-type and a *p*-type, and join them at their hot ends [*see illustration on the preceding page*]. Between their cold ends we place a conductor through which we wish to pass a current. This conductor may be the armature of an electric motor, a lamp, an electrolytic bath to reduce aluminum, or any other device using an electric current. Let us assume that a high temperature is maintained at the hot junction, and that the cold ends of the semiconductors are maintained at a lower temperature. The current produced in the *n*-type semiconductor flows from the hot to the cold end, while that in the *p*-type semiconductor flows from the cold end to the hot. The current thus flows around the whole circuit, including the electrical device. Such a thermoelectric cell, it is true, yields only 10ths of a volt, whereas technological applications require dozens and hundreds of volts, *e.g.*, the 100 to 200 volts used in the home. To obtain these voltages in a thermoelectric generator we need only join hundreds of individual thermoelectric cells together.

The quality of a thermoelectric cell, however, is not only determined by the voltage it will produce. Two other factors must be taken into account: its electrical and thermal conductivity. If the voltage it produces is to be delivered as useful current, then it must have high electrical conductivity. The same is true of any other generator: we want the electrical power outside, where we can use it. On the other hand, if a thermoelectric cell is to convert a high percentage of the heat energy into electrical energy, it must have low thermal conductivity. The principal deficiency of thermoelectric cells, as contrasted with other heat engines, is that most of the heat supplied to the hot end flows directly and wastefully, by heat conduction, to the cold end. Thus the ratio between the useful electrical output and the heat input in a thermoelectric cell is low.

These three factors—the thermoelectric power and the electrical and thermal conductivity—are inherent in the semiconductor material. Temperature is another important factor; the higher the temperature of a cell, the greater its electrical output. One of the prime objectives of current research is to develop materials with high inherent thermoelectric quality that will withstand high temperatures. The best materials developed to date convert heat to electricity with an efficiency of about 10 per cent

We may compare this to a steam-driven electric plant whose efficiency is about 30 per cent, or to a gasoline or kerosene engine with an efficiency upwards of 40 per cent. Clearly this comparison does not speak in favor of thermoelectric cells. But we have omitted the factor of time. We should recall that Seebeck's first thermoelectric cells had an efficiency of 3 per cent, just as high as the steam engines of their day. In the intervening century the efficiency of steam engines has been improved by a factor of 10, while thermoelectric technique—as represented by thermocouples made of metal—was permitted to drop backward by the same factor, to an efficiency of only .3 per cent. Semiconductors have now improved the situation somewhat. Thermoelectric cells today are not 100 times less efficient than steam engines, but only three times.

Is this the best we can do? Before agreeing that it is, let us consider the question from yet another aspect. Efficiency is an important characteristic of

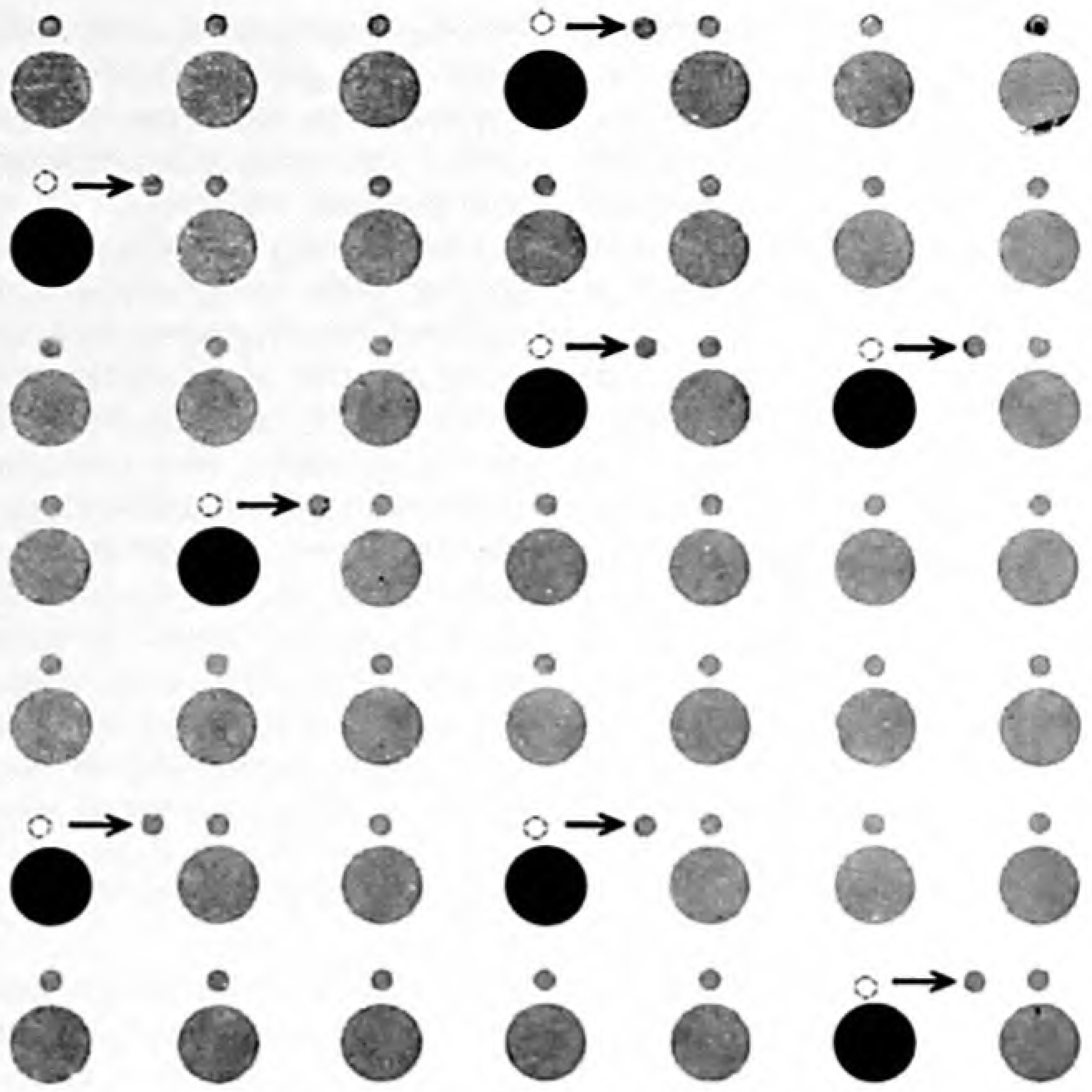
a machine, but it is not the only one. In order to obtain electrical energy from a steam engine, one must construct a furnace, a condenser, a steam boiler, a steam engine and a dynamo. This is complex and expensive equipment. A thermoelectric generator requires only a heater and a cooler; it has no moving parts. In many cases this advantage may more than compensate for lower efficiency, especially since an efficiency of 30 per cent can be obtained only from very powerful steam turbines. The efficiency of small steam engines may be as low as 10 per cent.

Thus for small power requirements, when one needs merely a few kilowatts of electricity, thermoelectric generators can compete with steam engines. For very low power requirements (as in radio, telegraph and telephone communications) thermoelectric generators provide the best engineering solution. And we must remember also that an efficiency of 10 per cent is not the limit for thermoelectric generators. The efficiency will increase significantly if one is able to go on to higher temperatures. If the temperature of the hot end could be raised to 600 degrees centigrade, for instance, the efficiency would go up to 18 per cent.

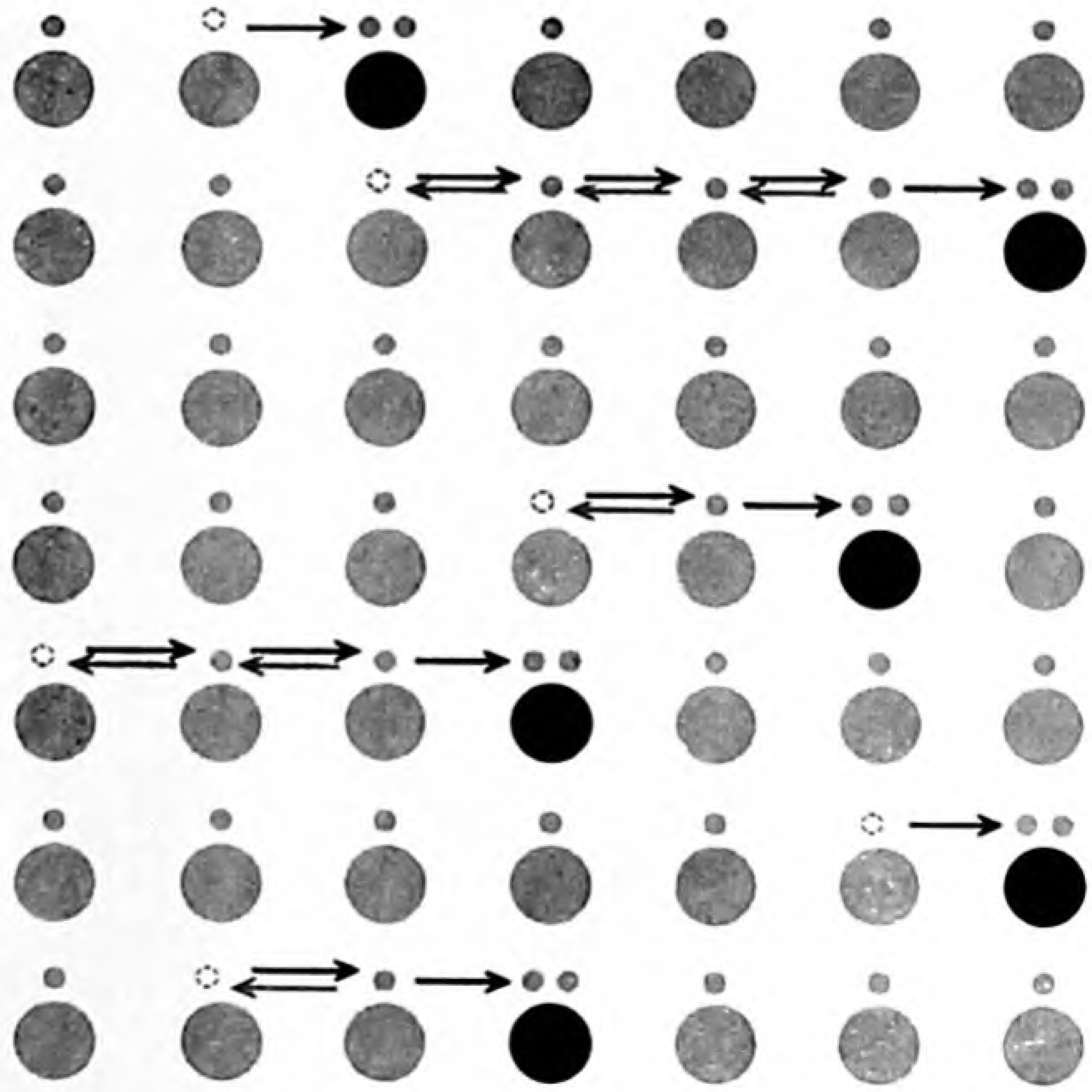
Even at their present efficiencies, however, thermoelectric generators are rendering effective practical service at many places that otherwise would be deprived of electric power. The thermoelectric generator shown in the illustration at left below obtains from the heat of an ordinary kerosene lamp enough electrical energy to power a radio receiving-set. Such thermoelectric lamps are being produced by tens of thousands in the U.S.S.R. and are in wide use.

Imagine someone living in the far north of our country, to whose dwelling such a lamp is brought. Although the snows and the tundra separate him from the rest of humanity, he is suddenly able to use a radio to hear the news of the day and music, and to learn about the life of his country. At our laboratory in Leningrad we receive moving letters from such hermits.

On the agricultural lands of the U.S.S.R. separate groups of farmers often work far apart. They are able to communicate with the farm headquarters by means of radio transmitters whose range is 30 miles or more. Such radio stations are supplied with electrical energy by thermoelectric generators which develop about 15 watts when heated by a kerosene stove. In isolated communities stoves heated by wood or



N-TYPE SEMICONDUCTOR is a crystalline material composed of an element like germanium (*gray balls*) which has tightly bound electrons (*small gray balls*), plus an "impurity" element (*black balls*) which has loosely bound electrons that are free to conduct electricity. The free electrons are shown moving toward the right in the direction of the potential gradient. Positive "holes" (*colored circles*) vacated by electrons are stationary.



P-TYPE SEMICONDUCTOR material contains an impurity element (*black balls*) with fewer electrons than it needs to satisfy the bonds to its neighboring atoms in the crystal. Electrons from the neighboring atoms move in to satisfy the deficiency, leaving positive holes, which are in turn filled by other electrons. Thus the holes can wander through the crystal. Under an applied voltage their direction of motion is opposite that of electrons.

some other local fuel can be used to generate from a few hundred watts to a full kilowatt for illumination or communication. At an output of one kilowatt this generator uses only 5 per cent of the heat supplied; the remaining 20 kilowatts or more of heat can be used to warm the pens of livestock.

And how much heat is expended uselessly! All heat engines give up more than half of their heat to the air, and this at temperatures at which thermoelectric generators could supply significant amounts of additional electrical energy. Let us recall, for instance, how hot the exhaust pipes of automobiles are. Anyone who lives in the modern technological world and is not without imagination can think of dozens of applications for semiconductor thermoelectric generators.

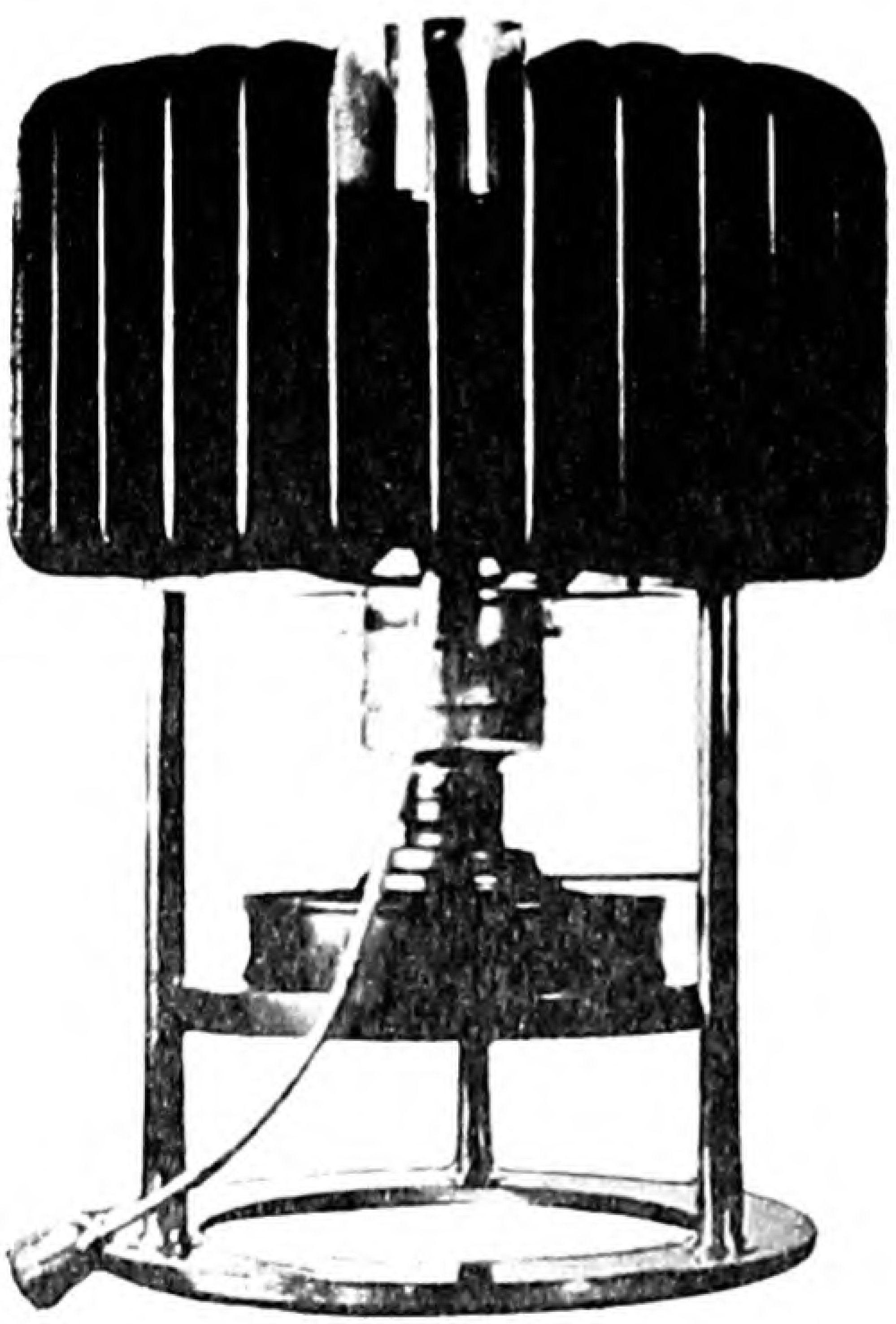
Of course our greatest source of heat now going to waste is the sun itself. Each square meter exposed to the sun receives power at the rate of about a kilowatt. Over the entire earth this adds up to about 100,000 billion kilowatts, a million times more than all of the electrical power produced throughout the world. In just a few days the sun supplies as much energy as can be recovered from all known coal and oil reserves, accumulated over billions of years. The plants responsible for this accumulation transform less than 1 per cent of the energy received from the sun into chemical reserves.

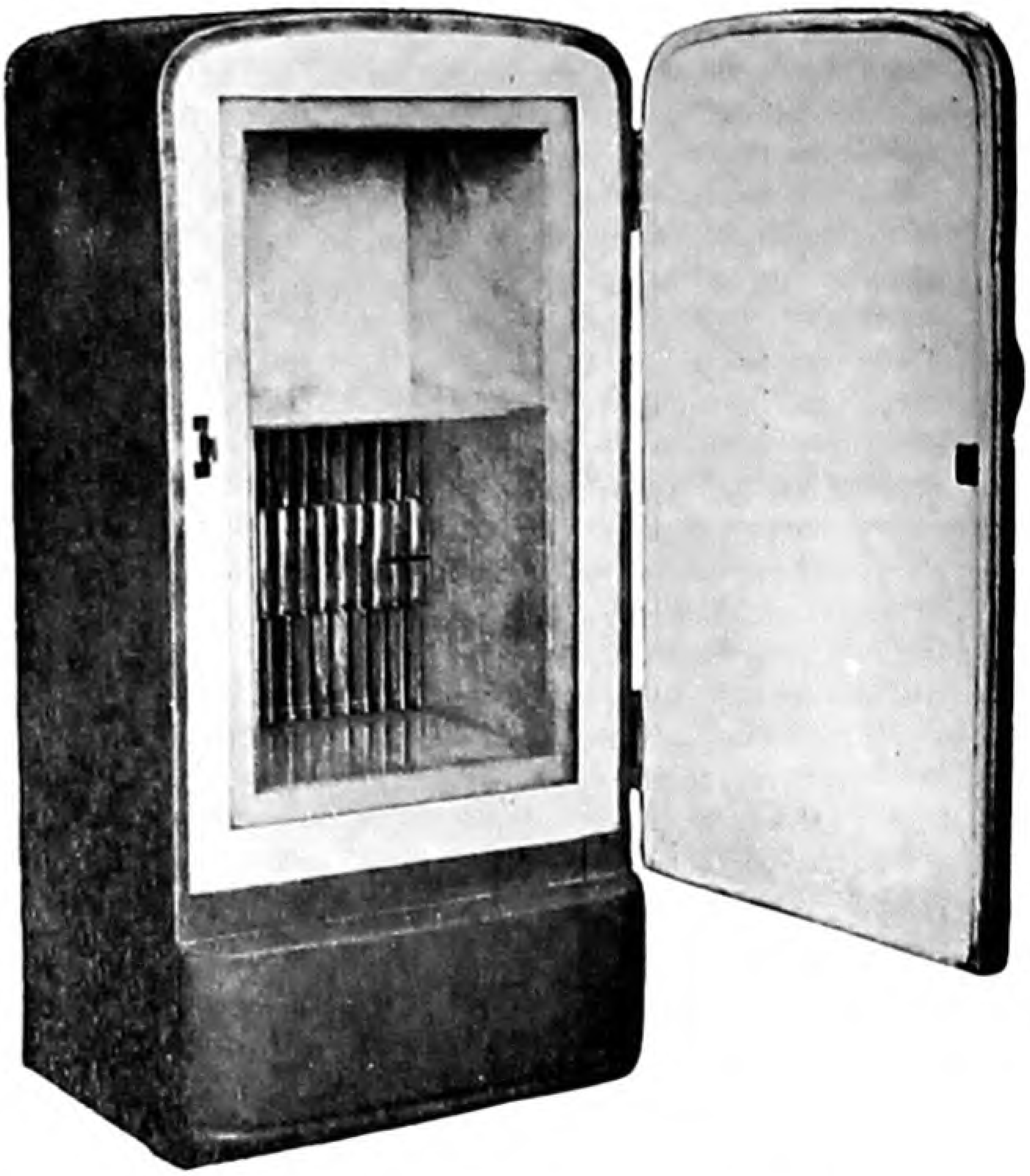
Solar energy indeed has its weak points. With 5 per cent efficiency, a million-kilowatt power plant would require an area of 10 square miles and would function only on sunny days and during daylight. The area necessary, however, is not a significant point; more important are the capital and maintenance costs. If a solar generator costs much more per kilowatt than a steam or hydroelectric plant, then solar generation of electricity is unprofitable. So far solar power plants, designed around conventional steam systems or around the most advanced solar batteries, have proved too expensive even on paper.

But let us now consider thermoelectric solar generators. Calculations and preliminary experiments indicate that small thermoelectric units are entirely feasible, even allowing for the cost of the large steerable mirrors necessary to concentrate the sunlight. Such units could be used to pump water from underground wells and irrigate desert land. This possibility strongly commends itself: irrigated deserts are the world's best gardens. It is still difficult to say anything about large thermoelectric solar plants. Yet one should never forget that a region with dimensions of the order of 100 kilometers could provide enough electric power for the whole earth. Perhaps thermoelectric cells will yet find their place in both the small- and large-scale exploitation of solar energy.

It must be pointed out, however, that the thermoelectric generation of electricity still faces many technical difficulties. The high temperatures necessary for good efficiency are harmful to semiconductor materials; oxidation reduces their thermoelectric quality and heat stresses tend to cause cracks. Low-temperature thermoelectric generators can be discussed with more confidence. A generator which we constructed 10 years ago, and which has been operated at room temperatures or lower, still has its original properties.

The transformation of heat into electricity is, of course, only half the story of the thermoelectric cell. It can transform energy the other way and, in doing so, serve the opposed functions of heating or cooling. To see how it does that, let us look again at the illustration on page 173. There, with the cell acting as





THERMOELECTRIC TECHNOLOGY IN U.S.S.R. has produced the devices shown on these two pages. The kerosene-lamp chimney at left employs the Seebeck effect to generate enough electric current to power a radio. The three devices at right employ the Peltier

a generator, the hot junction receives heat, while the cold gives off part of this heat. The remaining heat is converted to the electrical energy which produces the current flowing around the cell. What Peltier discovered is that the current which flows across the hot junction extracts heat from it and transfers this heat to the cold junction.

Now if we supply electrical energy instead of heat to the circuit and cause a current to flow through it, this current will necessarily extract heat from the hot junction and liberate heat at the cold junction. Thus the electric current will cool one junction and heat the other. This is the operating principle of thermoelectric refrigerators and heaters.

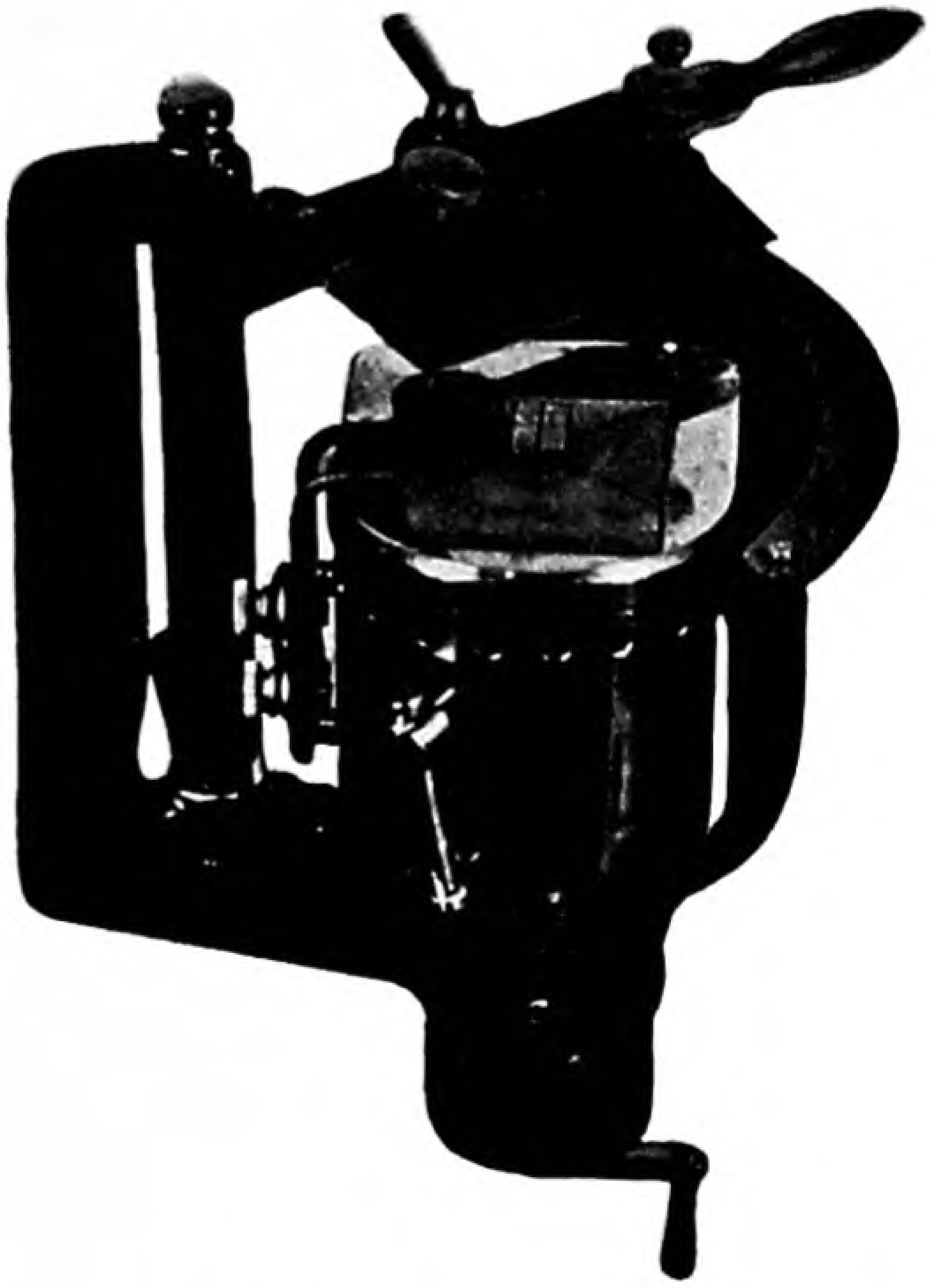
The thermoelectric refrigerators developed in our laboratories for domestic use have an efficiency approaching that of the more complicated and expensive refrigerators currently in use. Some of the conventional units use less electrical energy, but this is not particularly important, since the energy expended is not great. The cost of thermoelectric cells, moreover, runs considerably below the cost of electric motors and compressors.

The cost of the cell, however, is not the whole story. A bigger item of expense is the heat-exchange system needed to remove the heat from the refrigerator. In addition, since thermoelectric cells operate on direct current, a rectifier must be used where power is supplied by alternating current. As yet it is hard to predict what thermoelectric refrigerators for the home will cost when they are mass-produced. The prospects are, however, very hopeful.

When it comes to large-scale refrigeration, as in the food industry, the thermoelectric cell does not compete successfully with mechanical systems. Here the amount of energy consumed is far from immaterial, and modern machines use it with greater efficiency than any thermoelectric cell so far developed or in prospect. We must therefore assume that the applications of thermoelectricity to refrigeration will be restricted to those cases in which it is more important to avoid complex machinery than to keep down the amount of electrical energy expended.

There are many such cases. Although it is easy to obtain heat from an electric current, to use it for cooling has always required electric motors and compressors. But where is this bulky gear to be placed if, for instance, one would like to be able to vary the temperature of an object observed under a microscope? A small microscope stage provided with thermoelectric refrigerating cells requires only a few watts to lower the temperature of the specimen down to 50 degrees below zero centigrade, or, by changing the direction of the current, to raise it to 80 degrees above. This provides the biologist and the chemist with a much-needed facility. Another useful device is a rod that is cooled or heated by thermoelectric cells. As soon as a current is passed through it, frost appears on its surface, and in a minute or two the temperature drops to 30 degrees below zero. This rod can be used to cool any object, such as the skin of an animal or the air in a small box. Changing the direction of current quickly heats the rod, and water boils on its surface. If something has to be heated or cooled, it cannot be done more conveniently than by this device. The stronger the current, the greater the temperature change in both cases. Moreover, the temperature is subject to precise control, to within .001 degree.

We now have about 30 thermoelectric devices for various purposes, and they are all useful. It is hardly worth enumerating them. Rather I would call on the



effect for refrigeration. The domestic refrigerator second from left is the prototype of a mass-production unit. Third from left is a constant-temperature container for transporting biological specimens; at right is a microtome with a refrigerated specimen-holder.

ingenuity of the reader; he will undoubtedly be able to think up 30 other applications!

Instead, I should like to direct the reader's attention to still another aspect of the matter: namely, heating by means of thermoelectricity. What occurs is at first glance unexpected. In order to heat a room with an ordinary electric heater at the rate of a kilowatt, the heater must also use up electrical energy at the rate of a kilowatt. But if thermoelectric cells are used for this purpose, the expenditure of half a kilowatt or less will yield the same result!

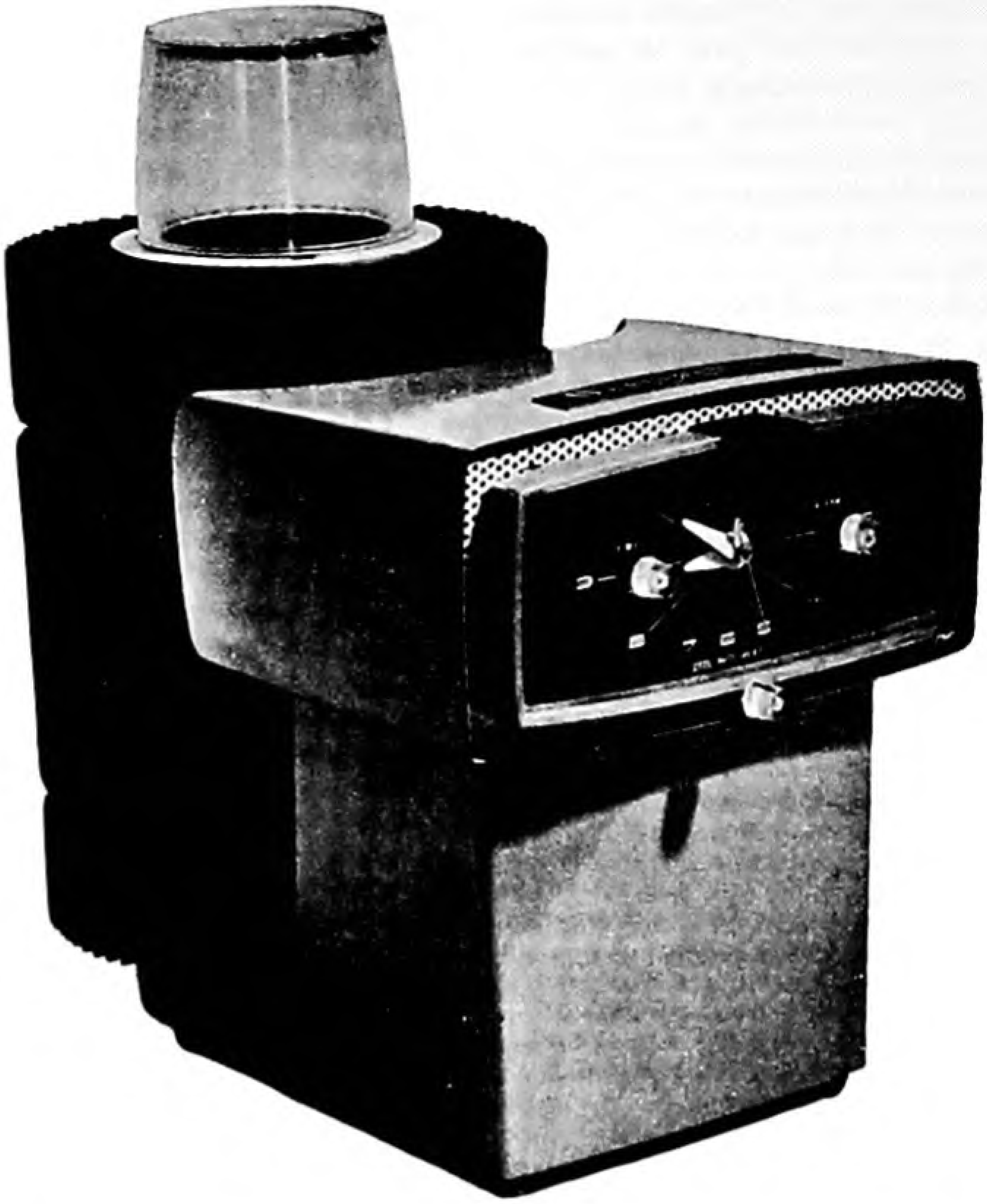
One might think that thermoelectric cells shake the very foundation of science: the law of conservation of energy. But of course this is not so.

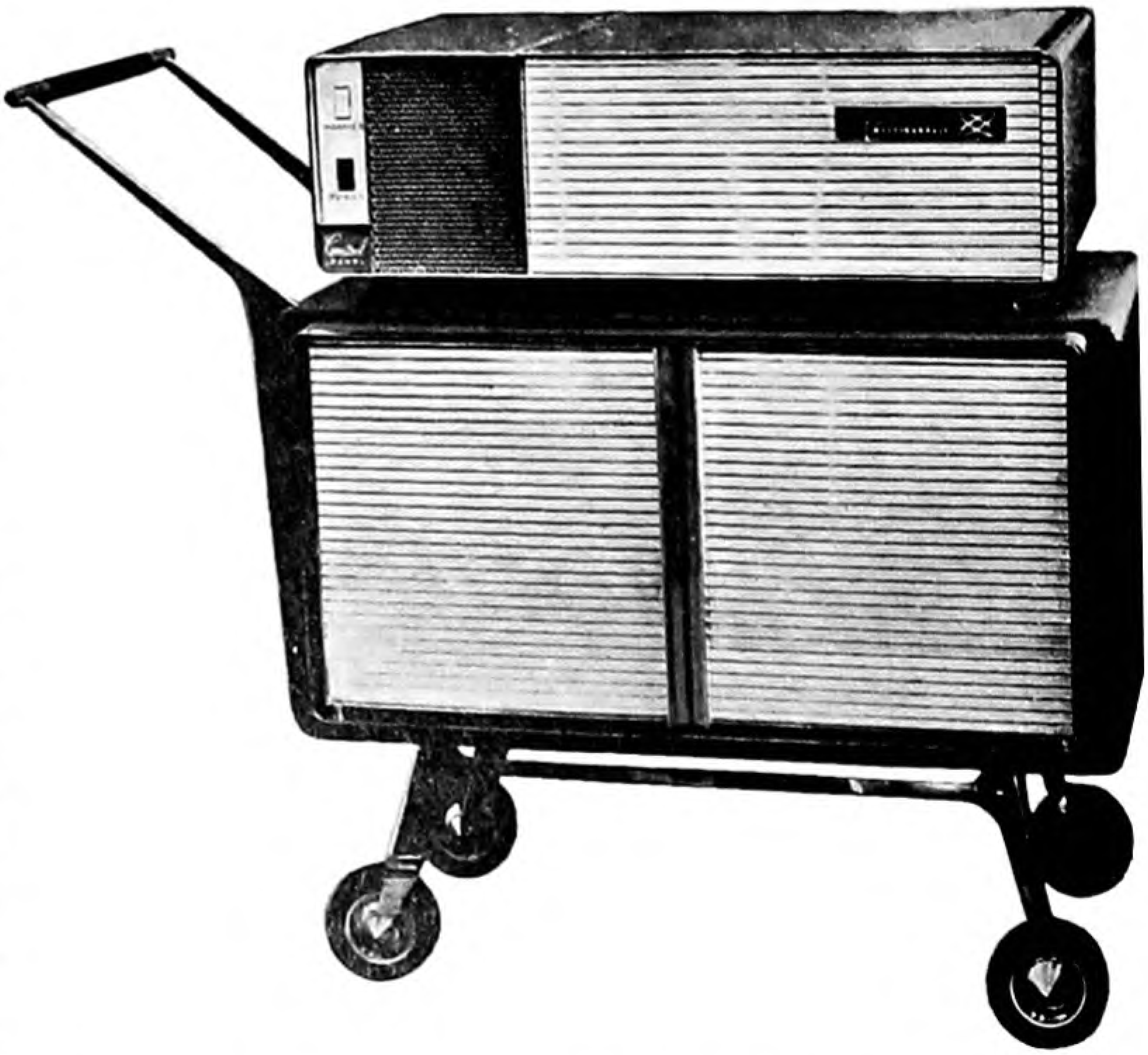
This extra heat supplied to the room is not created from nothing, but is transferred from a colder source, such as from the water supply. The thermoelectric cell, acting as a "heat pump," removes heat from the water, and transfers this heat to the room together with the heat supplied by the current. Even though the heat reservoir may be at a low temperature, its heat can be delivered at a higher temperature by means of thermoelectric cells. The less the temperature difference between the reservoir from which the heat is removed and the warm one to which it is supplied, the less electrical energy is needed. At a temperature difference of 10 degrees our thermoelectric cells can transfer heat with an expenditure of one fifth of a kilowatt for every kilowatt delivered; at a temperature difference of 20 degrees the energy expended goes up to one third of a kilowatt, and at 30 degrees to one half a kilowatt. Mechanical heat pumps used for this purpose have always been extremely complicated.

On this note we shall cease enumerating the devices which have already been developed and which may yet be developed from semiconductor thermoelectric cells.

If in disappointment the reader says, "Why nothing has yet been done," let him remember that the pace of modern ideas is constantly accelerating. Aviation and radio have grown rapidly and moved forward briskly. Cinema and television have unfolded even more quickly.

Now thermoelectricity is unfolding before our eyes. It is only in the last two or three years that this field has been opened. Let us see what will happen in the next three to five years!





THERMOELECTRIC TECHNOLOGY IN U. S. has produced these prototypes of products recently publicized by Westinghouse Electric Corporation. Clock-controlled bottle-warmer (and cooler) at top and hostess cart at bottom use Peltier effect for refrigeration and heating.

The Author

ABRAM F. JOFFE is a leading Soviet physicist and a member of Russia's central scientific body, the Academy of Sciences, in which he heads the Institute of Semiconductors. Born in 1880, Joffe studied at the Technological Institute in Leningrad (then St. Petersburg) and the University of Munich. For many years he worked in collaboration with Wilhelm Konrad Röntgen, the discoverer of X-rays. Later he headed the Leningrad Physical-Technical Institute and the State X-Ray and Radiology Institute.

Bibliography

THE JUNCTION TRANSISTOR. Morgan Sparks in *Scientific American*, Vol. 187, No. 1, pages 28-32; July, 1952.

SEMICONDUCTOR ELEMENTS AND THERMOELECTRIC COOLING. A. F. Joffe. Infosearch Limited, 1957.

SCIENTIFIC AMERICAN May 1959, Vol. 200, No. 5, pp. 84-99 OFFPRINT 223

EXPERIMENTS IN COLOR VISION

by Edwin H. Land

The eye has recently been found to be an instrument of wonderful and unsuspected versatility. It can perceive full color in images which, according to classical theories, should be monochromatic.

From childhood onward we enjoy the richness of color in the world around us, fascinated by the questions: "How do we see color? How do you know you see the same color I do? Why do colors sometimes mix to give quite different colors?" Since 1660, when Isaac Newton discovered the properties of the visible spectrum, we have slowly been learning the answers; and we are finding that the beauty of the outer world is fully matched by the technical beauty of the mechanisms whereby the eye sees color.

No student of color vision can fail to be awed by the sensitive discernment with which the eye responds to the variety of stimuli it receives. Recently my colleagues and I have learned that this mechanism is far more wonderful than had been thought. The eye makes distinctions of amazing subtlety. It does not need nearly so much information as actually flows to it from the everyday world. It can build colored worlds of its own out of informative materials that have always been supposed to be inherently drab and colorless.

Perhaps the best way to begin the story is to consider two sets of experiments. The first is the great original work of Newton, which set the stage for virtually all research in color vision since that time. The second is an apparently trivial modification that reverses some of his basic conclusions.

As is so often the case with truly revolutionary insights, the simplicity of Newton's discovery causes one to wonder why no one before him had made it. He passed a narrow beam of sunlight through a prism and found that it fanned out into the band of colors we know as the visible spectrum: red, orange, yellow, green, blue, indigo and violet. When he reversed the process, gathering the beam together with a second prism, the colors vanished and white light reappeared. Next he tried recombining only parts of the spectrum, inserting a slotted board to cut off all but certain selected bands [*see diagram on page 184*]. When he combined two bands of color, letting the rays mix on a screen, a third color appeared, generally one matching a color lying between the bands in the spectrum.

Let us repeat this last experiment, placing the openings in the board just inside the ends of the narrow yellow band in the spectrum. When these two yellow beams strike the screen, they combine, as Newton observed, to produce yellow.

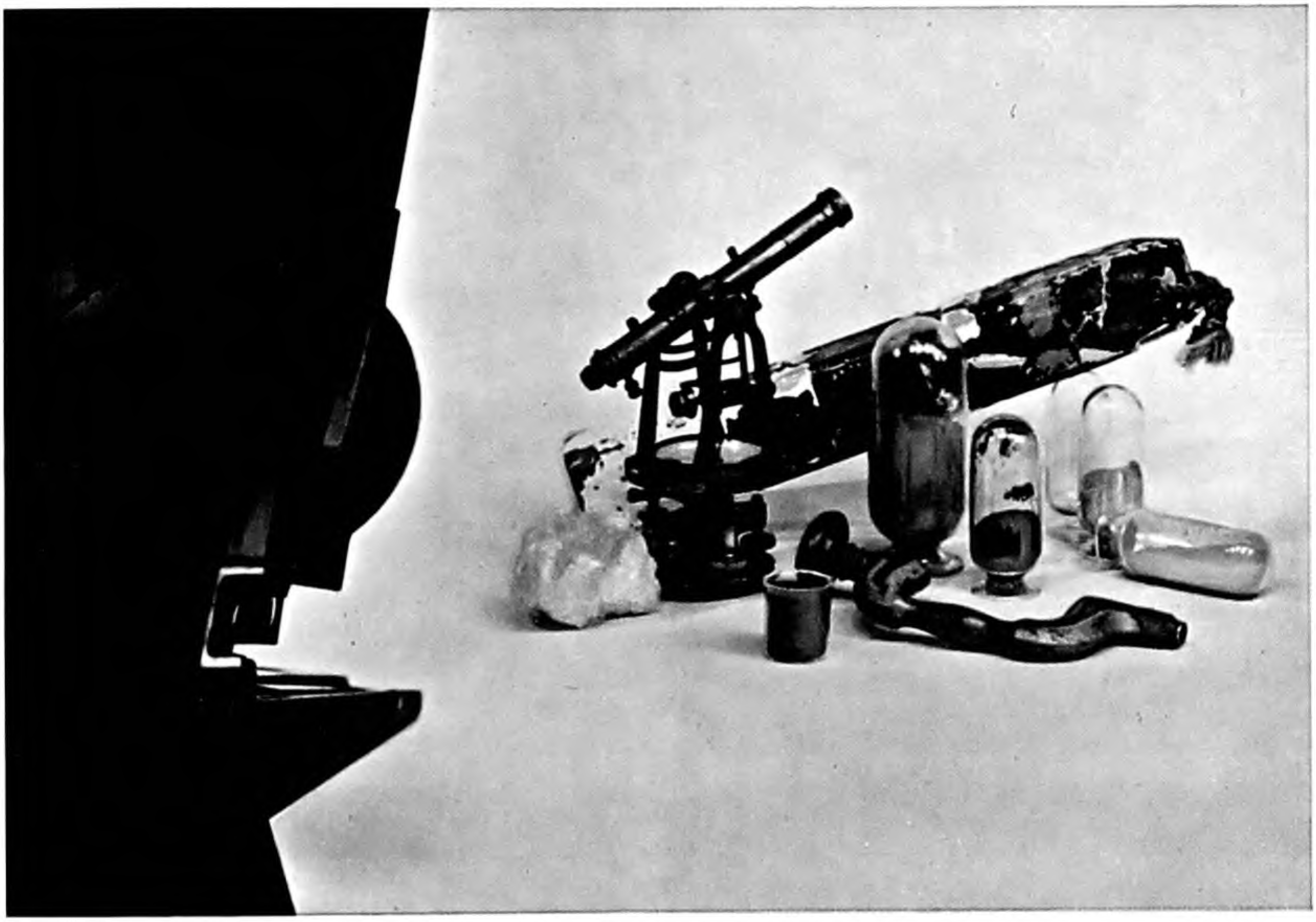
Now for our modification. In front of the slits we place a pair of black-and-white photographic transparencies. Each shows the same scene: a collection of variously colored objects. There is, of course, no color in the photographs. There are simply lighter and darker areas, formed by black silver grains on transparent celluloid. A glance at the two shows that they are not absolutely identical. Some of the objects in the scene are represented by areas which are lighter in the first photograph than in the second. Others are darker in the first and lighter in the second. But all that either photograph can do is to pass more or less of the light falling on its different regions.

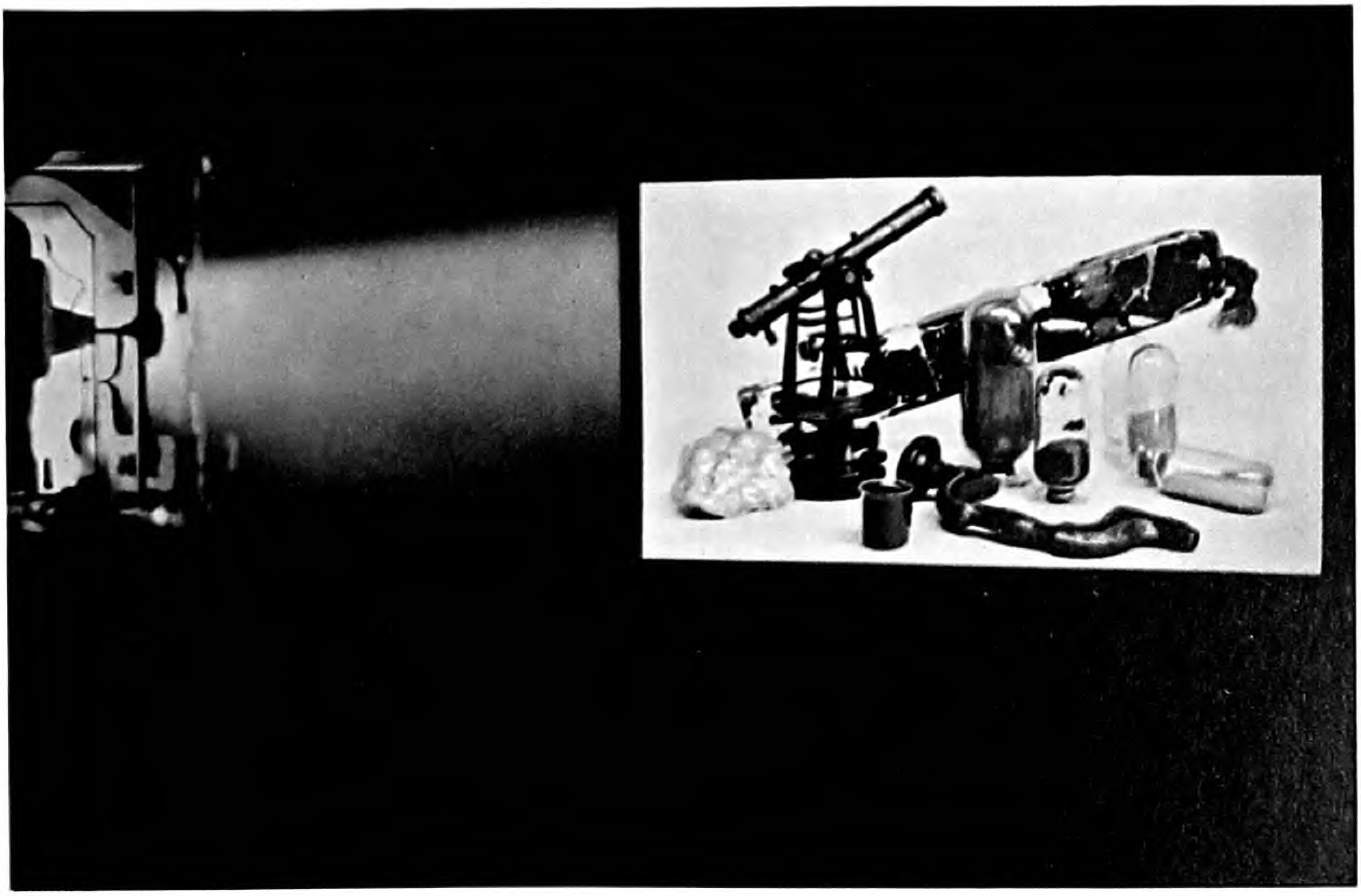
The yellow beams pass through these transparencies and fall on the screen. But now they are not yellow! Somehow, when they are combined in an image, they are no longer restricted to producing their spectral color. On the screen we see a group of objects whose colors, though pale and unsaturated, are distinctly red, gray, yellow, orange, green, blue, black, brown and white [*see bottom photograph on page 182*]. In this experiment we are forced to the astonishing conclusion that the rays are not in themselves color-making. Rather they are bearers of information that the eye uses to assign appropriate colors to various objects in an image.

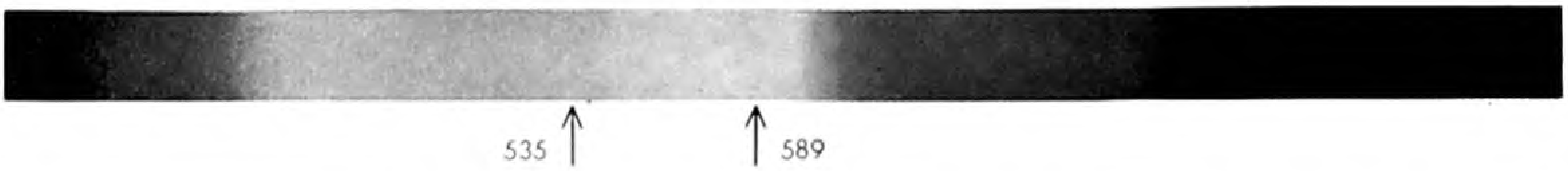
The Old Theory

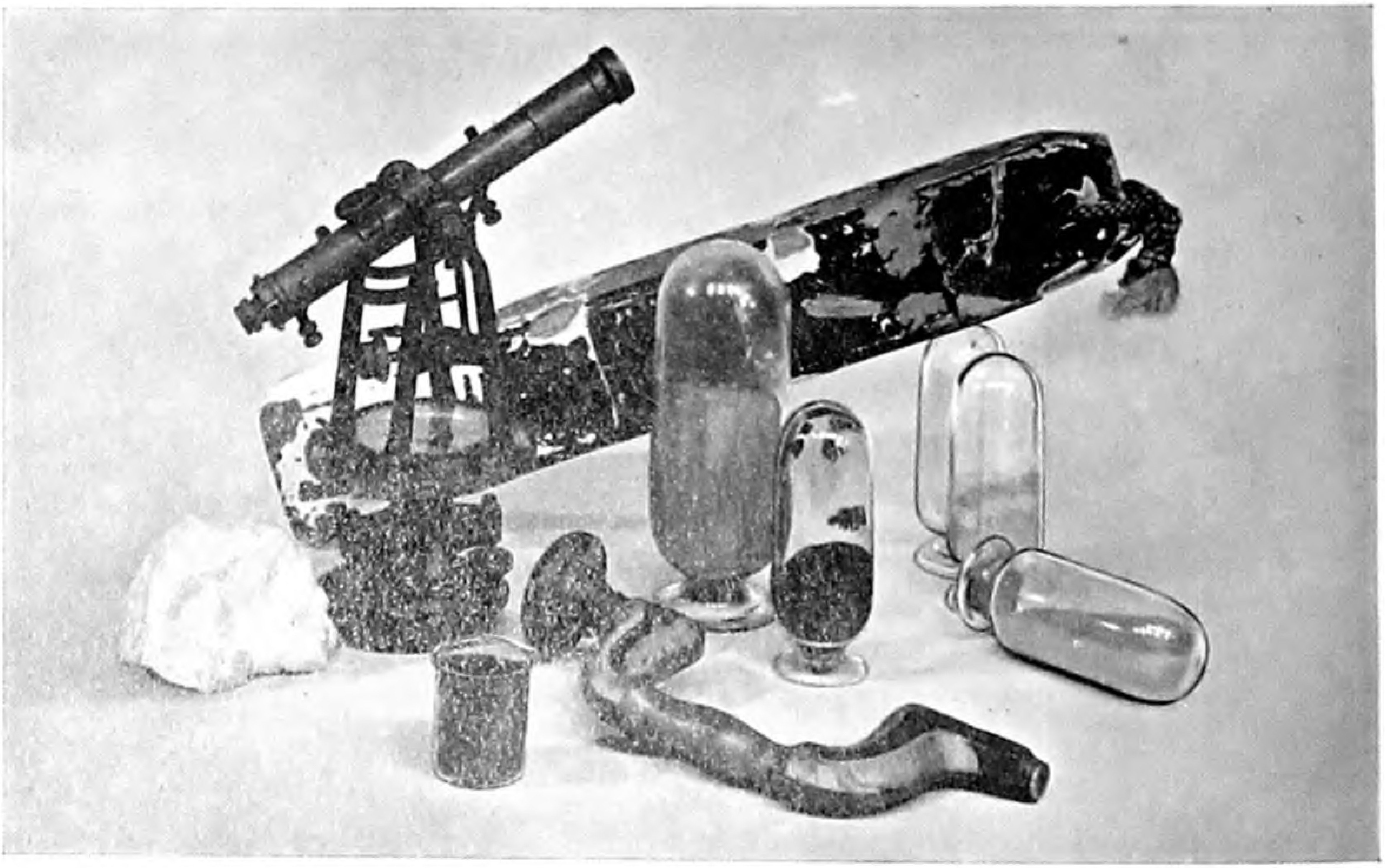
This conclusion is diametrically opposed to the main line of development of color theory, which flows from Newton's experiments. He and his successors, notably Thomas Young, James Clerk Maxwell and Hermann von Helmholtz, were fascinated by the problem of simple colors and the sensations that could be produced by compounding them. Newton himself developed quite good rules for predicting the colors that would be seen when various spectral rays were mixed to form a spot of light on a screen. These rules can be summarized in geometrical diagrams, one of the oldest of which is the color triangle [*see diagram at top of page 192*]. On new versions of it we can read off the result of com-

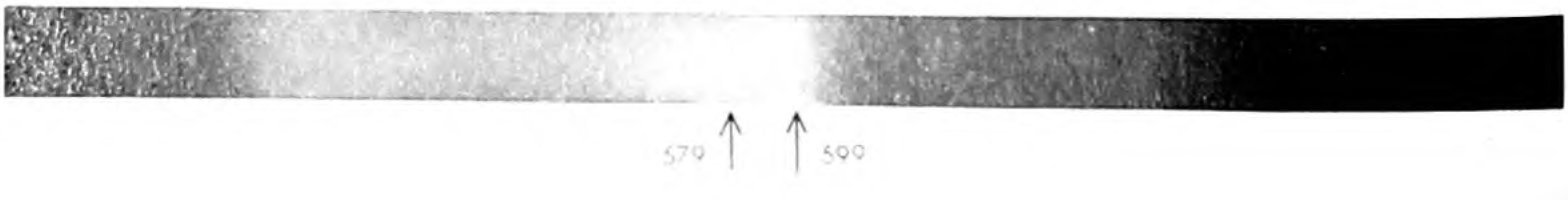
COLORED OBJECTS in the top picture on the opposite page were photographed with the special dual camera which appears at left. Here the two ground-glass screens of the camera are left uncovered to show that one image is photographed through a green filter and the other through a red filter. The images are photographed on ordinary black-and-white film; then black-and-white positive transparencies are made from the negatives. In the bottom photograph the "red" transparency is projected through a red filter and the "green" without a filter. When the two images are superimposed on the screen at right, they reproduce the objects in a full range of color.

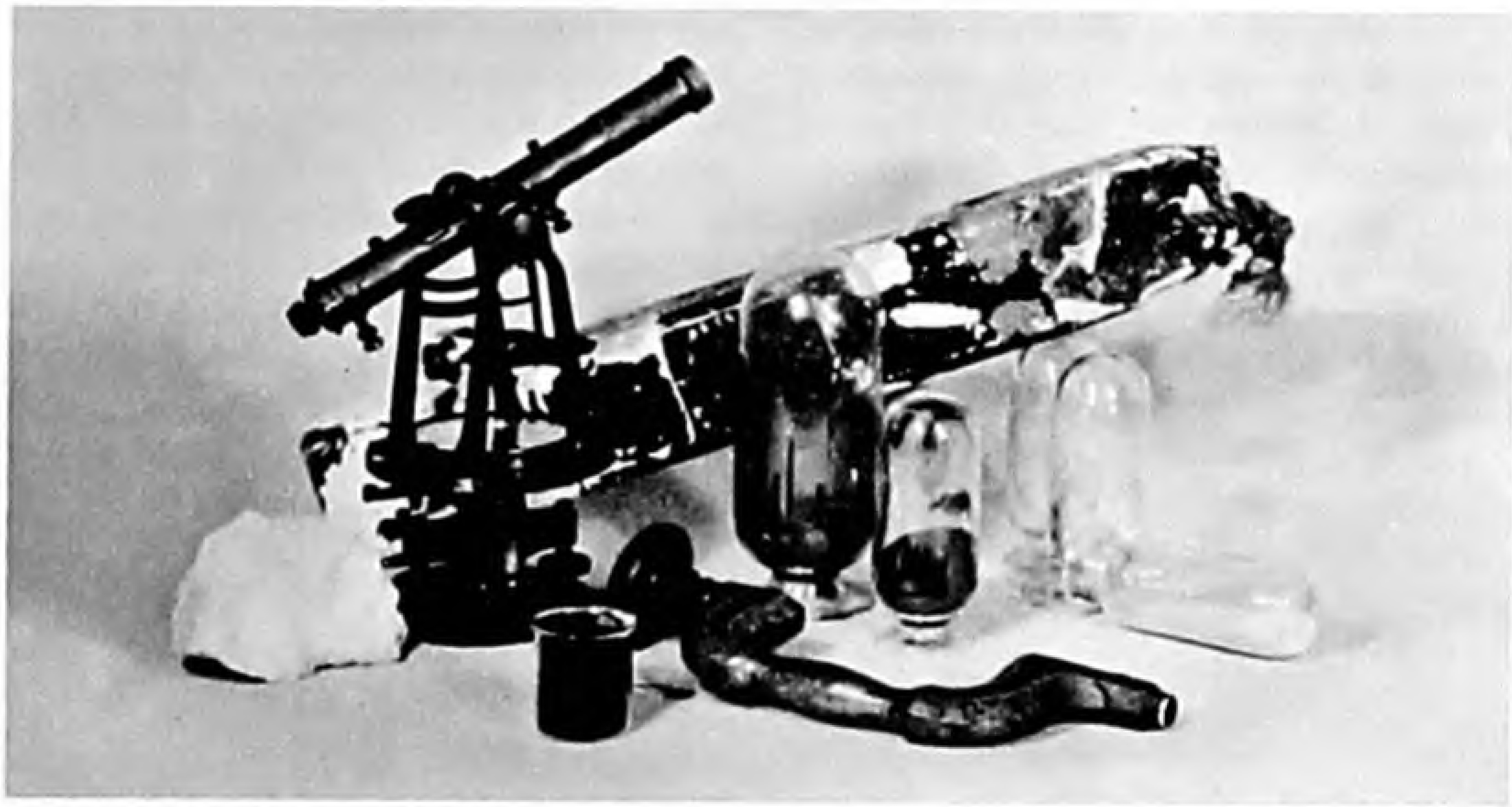





535
589




570
590



LONG AND SHORT RECORDS are provided by transparencies of these black-and-white photographs made through a red filter (*top*) and a green filter (*bottom*). In projection the long record (*top*) is illuminated by the longer of two wavelengths or bands of wavelengths, and the short record is illuminated by the shorter wavelength or band of wavelengths.

bining so many parts of color A with so many of color B.

Once it was discovered that light is a wave motion, the classical investigations of color acquired a deeply satisfying logical basis. The order of colors in the spectrum follows wavelength, the longest visible wavelength falling at the red end of the spectrum and the shortest at the violet end. A pure color would be a single wavelength; compound colors would be mixtures of pure colors.

In trying to match colors by mixing spectral stimuli Maxwell and Helmholtz found that three different wavelengths were enough to effect all matches, and that those wavelengths had to be chosen from the red, green and blue bands of the spectrum. Accordingly red, green and blue came to be called the primary colors. On the basis of this evidence they proposed a three-color theory of color vision. We need not go into the details here. The central idea is that the eye responds to three different kinds of vibration, and that all color sensation is the result of stimulating the three responses in varying degrees of strength. Thus it has become an article of faith in standard theory that the color seen at any point in a field of view depends on what wavelengths are issuing from that point and upon their relative strengths or intensities.

Now, as we have seen in our modification of Newton's experiment, the light at any point on the screen was composed of only two "yellow" wavelengths, yet the image was fully colored. And, as we shall see later, the colors in images will be remarkably stable even when the over-all relative strengths or intensities of the two wavelengths are varied.

Natural Images

Is something "wrong" with classical theory? This long line of great investigators cannot have been mistaken. The answer is that their work had very little to do with color as we normally see it. They dealt with spots of light, and particularly with pairs of spots, trying to match one to another. The conclusions they reached were then tacitly assumed to apply to all of color sensation. This assumption runs very deep, and has permeated all our teaching, except for that of a few investigators like E. Hering, C. Hess and the contemporary workers Dorothea Jameson and Leo M. Hurvich (who have studied the effect produced on a colored spot by a colored surround).

The study of color vision under natural conditions in complete images (as opposed to spots in surrounds) is thus an unexplored territory. We have been working in this territory—the natural-image situation, as we call it—for the past five years. In the rest of this article I shall describe some of the surprises we have encountered.

To form the image in our modification of Newton's experiment we needed two sets of elements: a pair of different photographs of the same scene, and a pair of different wavelengths for illuminating them. It is possible to make the pictures different by tinkering in the laboratory, arbitrarily varying the darkness of their different areas. But, as every photographer will have recognized at this point, a simple way to produce the two pictures is to make "color separations", that is, to photograph the scene through two filters that pass different bands of wavelengths. In this way the film is systematically exposed to longer wavelengths coming from the scene in one case, and to shorter wavelengths in the other. In our investigations we usually use a red filter for the longer wavelengths and a green filter for the shorter.

Now when we illuminate the transparencies with practically any pair of wavelengths and superimpose the images, we obtain a colored image. If we send the longer of the two through the long-wave photograph and the shorter through the short-wave photograph, we

COLORS SEEN when long and short records are illuminated by closely spaced narrow bands of wavelengths are reproduced in these "Flexichrome" photographs. The illuminating wavelengths are indicated by arrows on the spectrum under each photograph. These images could not be photographed directly; the response of color film to the limited range of wavelengths used here is very different from that of the eye. SCIENTIFIC AMERICAN has artificially reproduced the colors seen by the eye by adjusting the color in a Flexichrome print.

obtain most or all of the colors in the original scene and in their proper places. If we reverse the process, the colors reverse, reds showing up as blue-greens and so on.

Long Wavelengths *v.* Short

It appears, therefore, that colors in images arise not from the choice of wavelength but from the interplay of longer and shorter wavelengths over the entire scene. Let us now test this preliminary hypothesis by some further experiments.

There are several more convenient ways to combine images than in the arrangement of Newton's experiment. One of the simplest is to place the transparencies in two ordinary projectors, using filters to determine the illuminating wavelengths. The color photographs on pages 181 and 182 show images formed in this way.

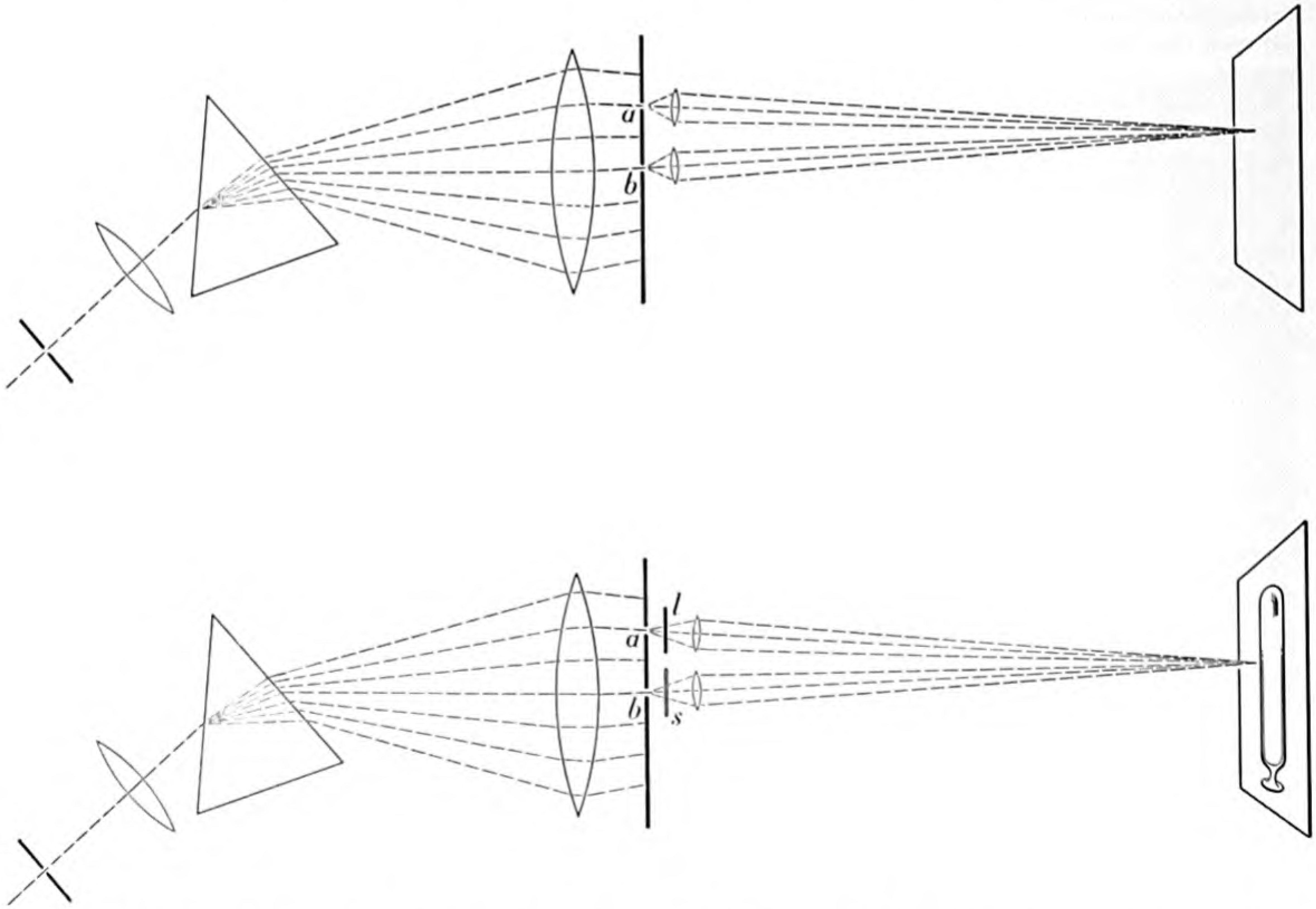
When we work with filters, we are not using single wavelengths, but rather bands of wavelengths; the bands have more or less width depending on the characteristics of each filter. It turns out that the width of the band makes little difference. The only requirement is that the long-wavelength photograph, or, as we call it, the "long record," should be illuminated by the longer band and the "short record" by the shorter band. Indeed, one of the bands may be as wide as the entire visible spectrum. In other words, it may be white light. The photograph on page 182, and the lower photograph on page 181, show the result of using a red filter for the long record and no filter (that is, white light) for the short record.

One advantage of this arrangement is that an observer can test the truth of our hypothesis in a simple and dramatic way. According to classical theory the combination of red and white can result in nothing but pink. With no photograph in either projector, and with a red filter held in front of one of them, the screen is indeed pink. Now the transparencies are dropped into place and the view changes instantly to one of full, vivid color. If the red filter is taken away, the color disappears and we see a black-and-white picture. When the filter is put back, the colors spring forth again.

An incidental advantage in using red for the long record and white for the short lies in the fact that the colors produced look about the same to color film as they do to the eye. Thus the image can be photographed directly. With more restricted bands of wavelengths the film, which does not have the new-found versatility of the eye, cannot respond as the eye does, and reproductions must be prepared artificially [*see photographs on page 182*].

The projectors afford a simple way of testing another variable: brightness. By placing polarizing filters in front of the projector's lenses we can vary the amount of light reaching the screen from each source. With no transparencies in the projector, but with the red filter still over one lens, the screen displays a full range of pinks, from red to white, as the strengths of the two beams are changed. When the photographs are in place, the colors of the image on the screen hold fast over a very considerable range of relative intensities.

Let us pause for a moment to consider the implications of this last demonstration. Remember that the photographs

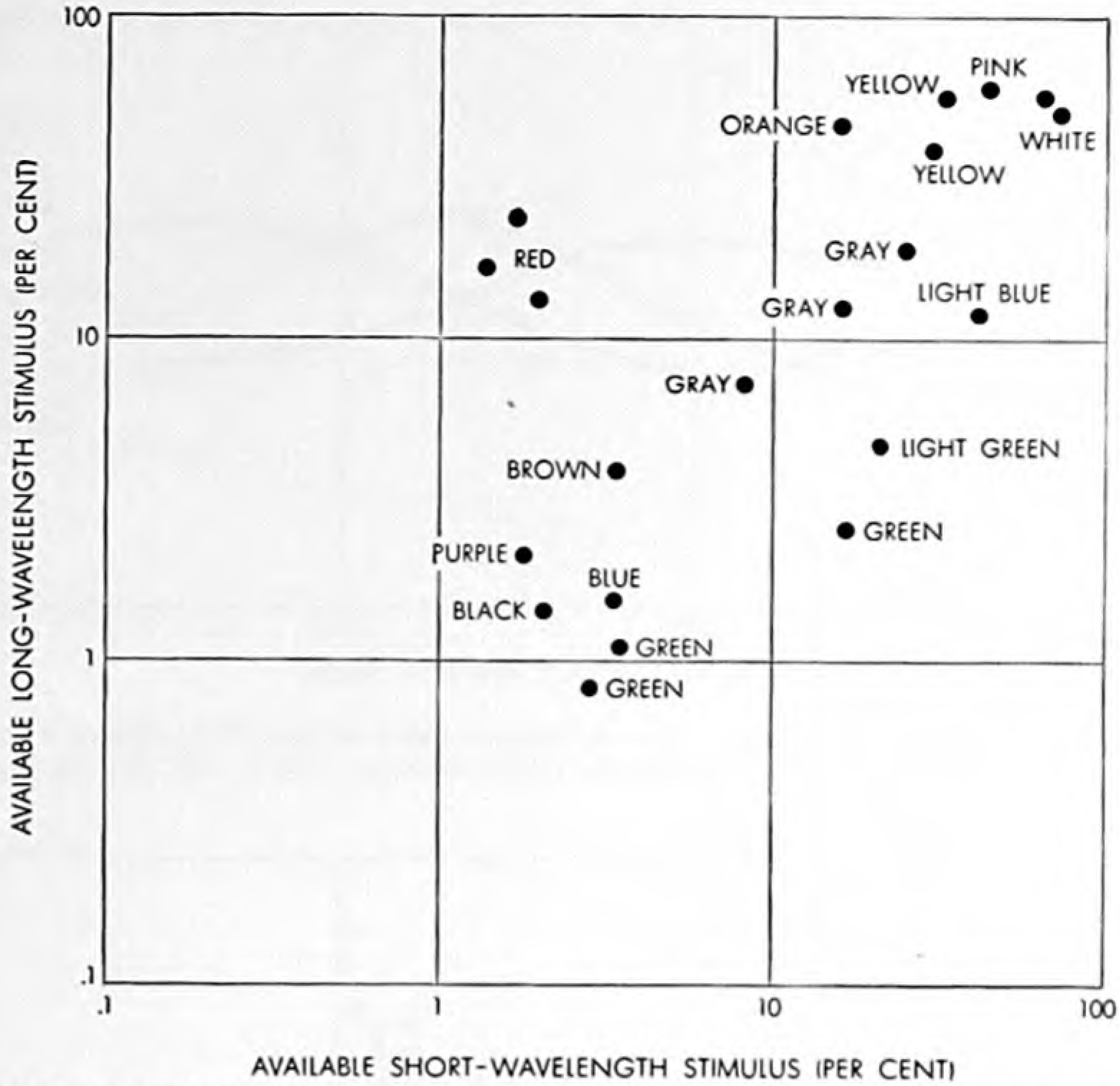


NEWTON'S EXPERIMENT in mixing spectral colors is shown schematically at top; the author's modification of the experiment, in which a pair of black-and-white transparencies is inserted in the beams, is diagrammed at bottom. When slits *a* and *b* are both in the yellow band of the spectrum, Newton's arrangement produces a spot of yellow on the screen. The image at bottom contains a gamut of color. The letters *l* and *s* in this diagram and others in this article refer, respectively, to the long record and the short record.

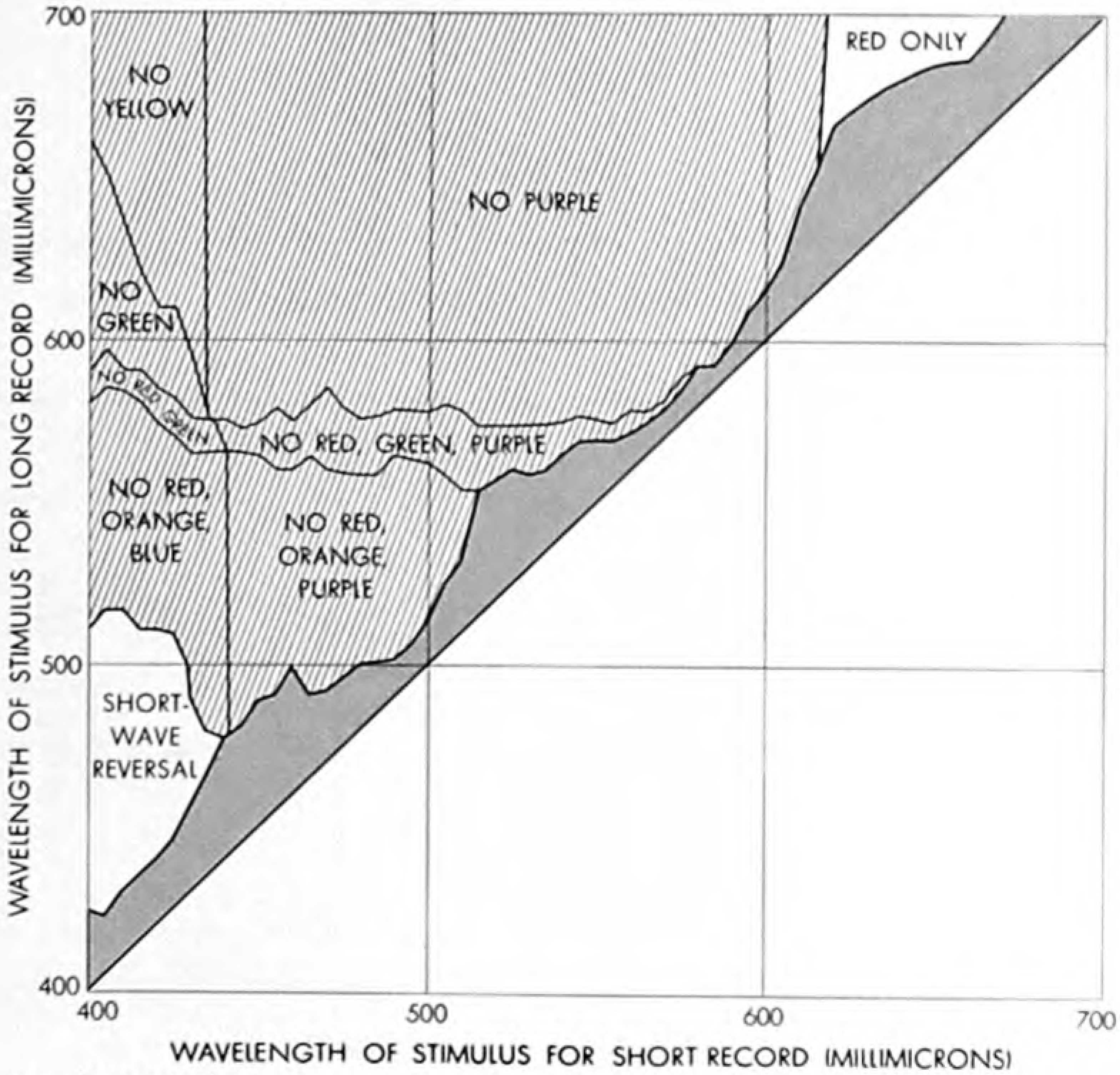
are nothing but pieces of celluloid treated to pass more light in some places than in others. All they can do to the red and white beams is to change relative intensities from point to point. In doing so they stimulate a complete gamut of color. Yet when we vary the relative intensities of the beams over the whole field of view, the colors stay constant. Evidently, even though the eye needs different brightness ratios, distributed over various parts of the image, to perceive color, the ratios that the eye is interested in are not simple arithmetic ones. Somehow they involve the entire field of view. Just how they involve it we shall see a little later.

The dual-projector system is convenient, but it is not a precision instrument. The wavelengths it can provide are limited by the characteristics of available filters. Narrow band-width filters may be used, but they seriously restrict the quantity of light. My colleague David Grey has therefore designed for me a dual image-illuminating monochromator [*see illustration on page 190*]. This instrument contains a pair of spectroscopes which allow us to light our transparencies with bands as narrow as we choose and of precisely known wavelength. By blocking off the spectroscopes and using filters, we can also obtain white light or broad bands. The two images are combined by means of a small, semitransparent mirror; light from one record passes through the mirror, and light from the other is reflected from its top surface. The intensity of each light source can be closely controlled.

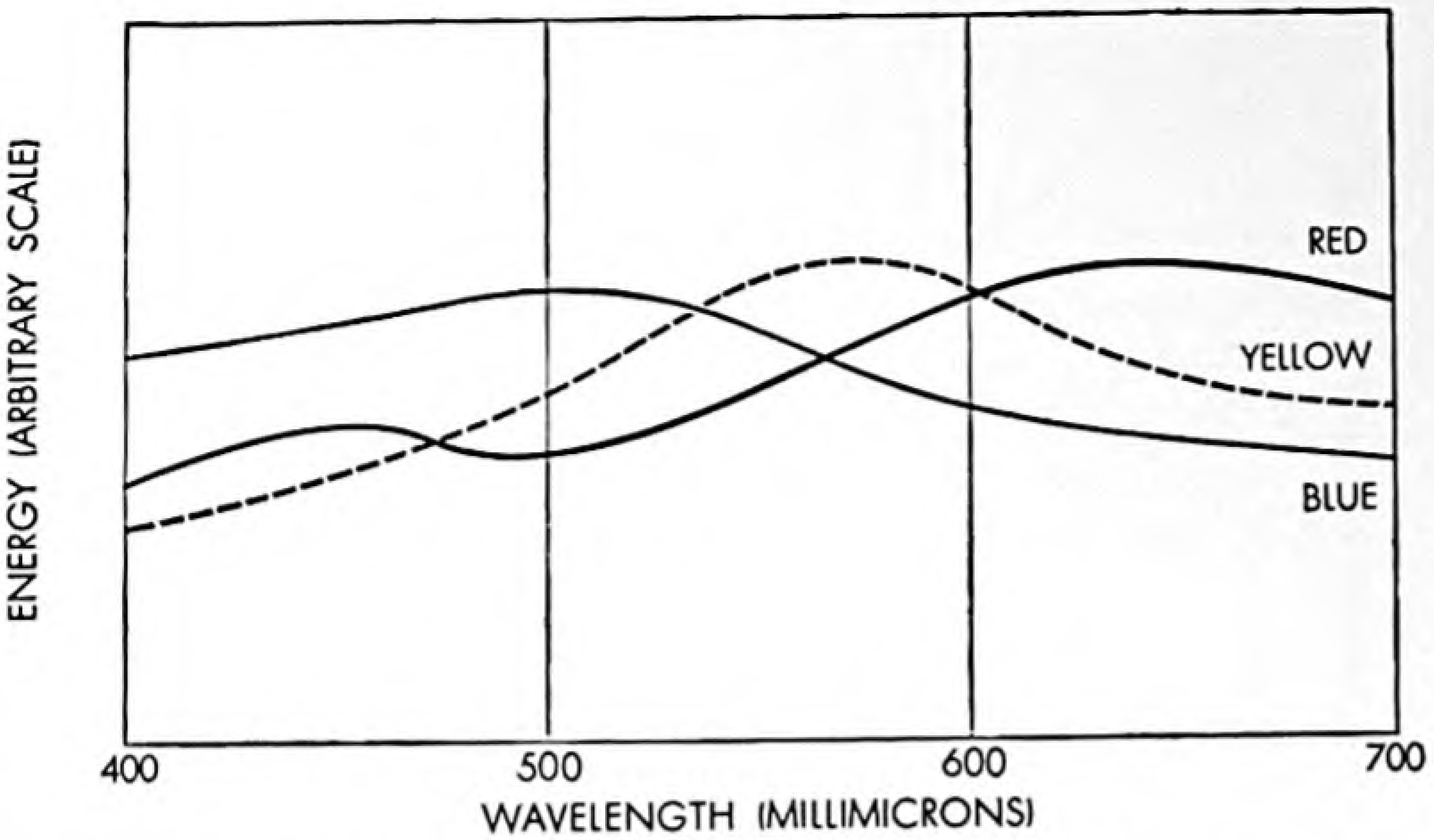
With the dual monochromator we have confirmed our broad hypothesis: Color in natural images depends on a varying balance between longer and shorter wavelengths over the visual field. We have also been able to mark out the limits within which color vision operates. It turns out that there must be a certain minimum separation between the long-record wavelength and the short. This minimum is different for different parts of the spectrum. Any pair of wavelengths that are far enough apart (and the minimum distance is astonishingly small) will produce grays and white, as well as a gamut of colors extending well beyond that expected classically from the stimulating wavelengths. Many combinations of wavelengths produce the full gamut of spectral colors, plus the nonspectral color sensations such as brown and purple. All this information has been summarized in a color map showing the limitations on the sensations produced by different pairs of wave-



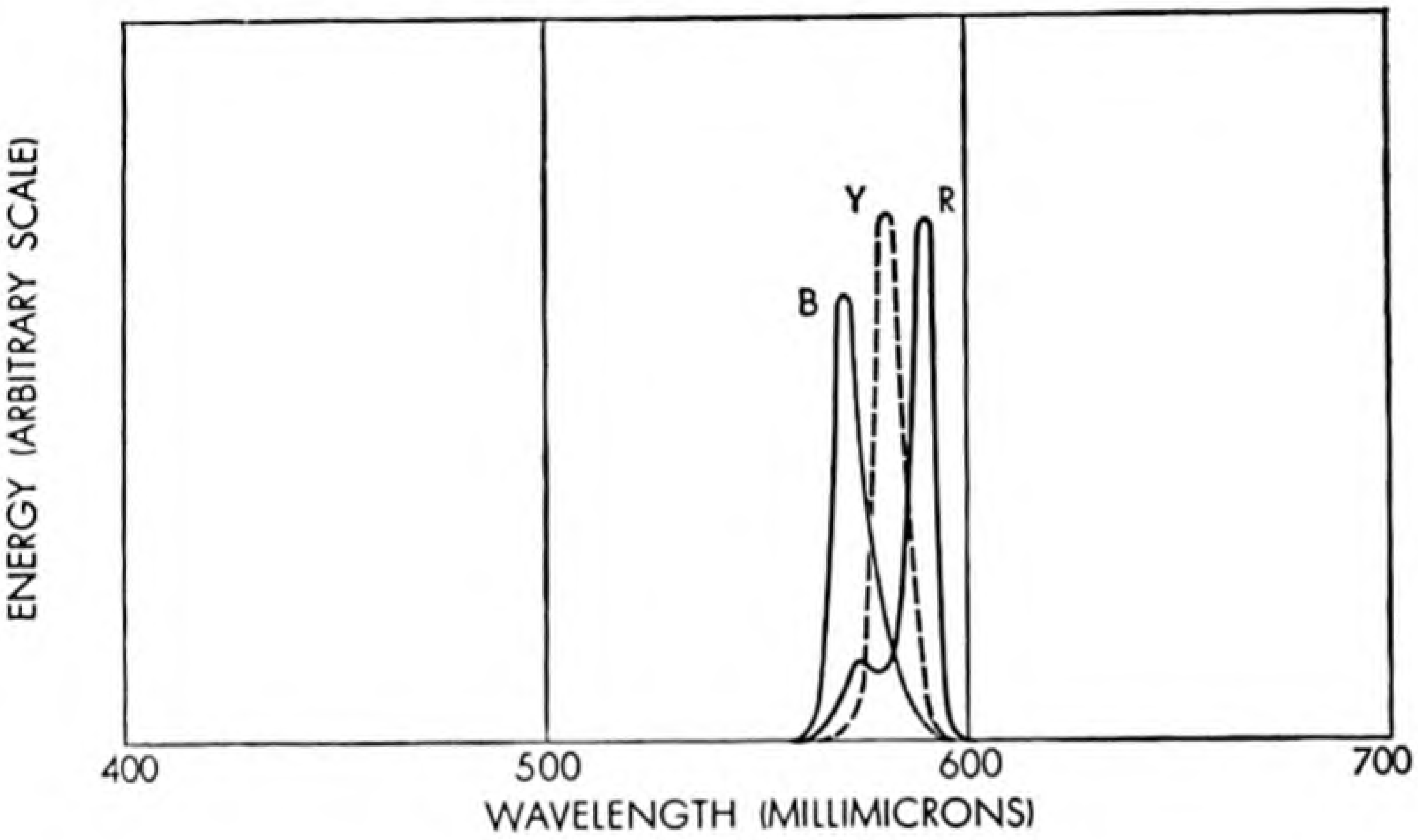
COORDINATE SYSTEM predicts colors in natural images. Axes are dimensionless, each measuring illumination at every point as a percentage of the maximum that could be there.

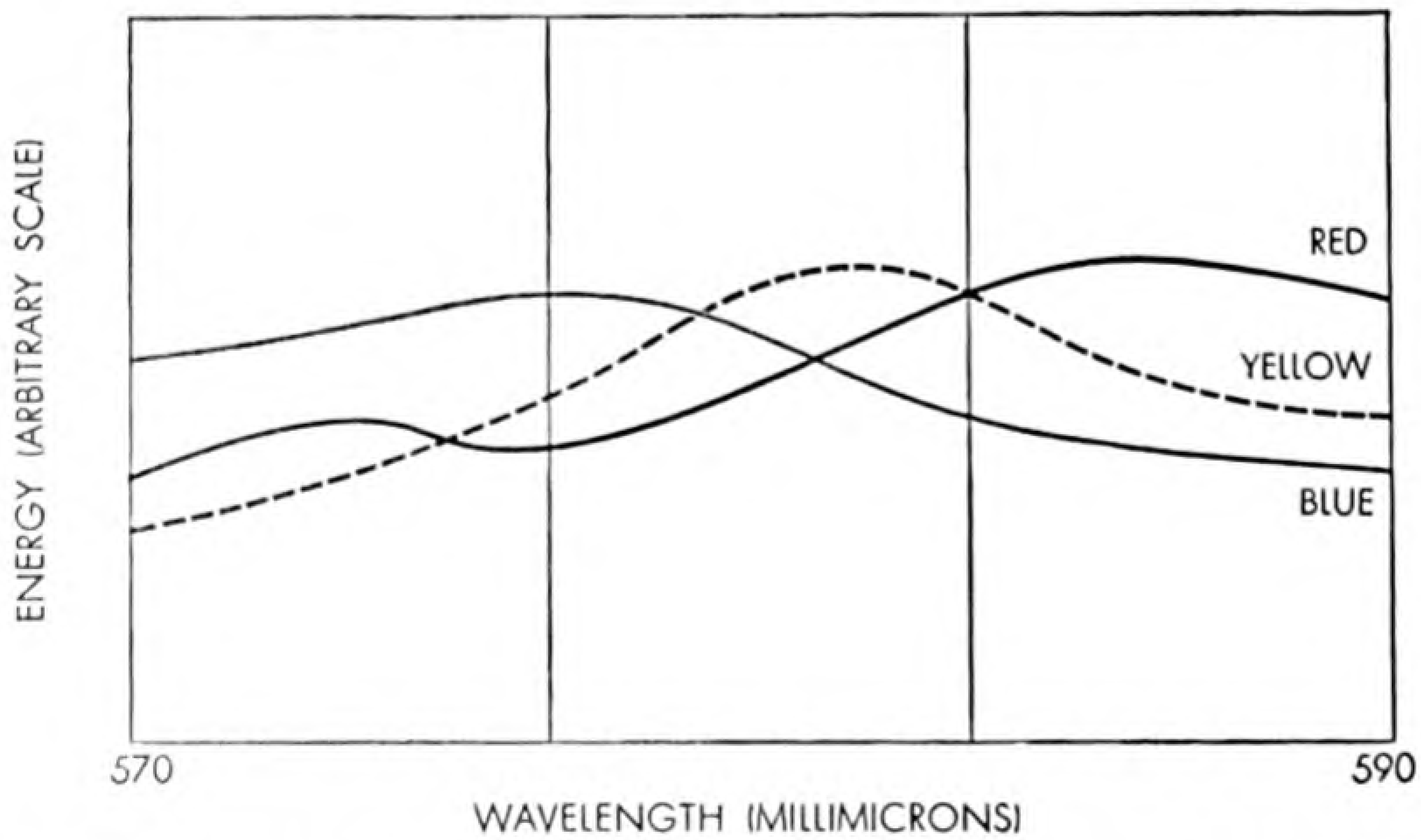


COLOR MAP shows limits on color obtainable with different pairs of wavelengths. The gray area is an achromatic region in which wavelengths are too close together to produce any kind of color. In the region marked "short-wave reversal" the colors are normal, but the short wavelengths act as the stimulus for the long record and the long wavelengths as the stimulus for the short record. The blank area below the diagonal is a region of reversed color obtained by illuminating the short record with long wavelengths and *vice versa*.



PIGMENTS IN OUR WORLD have broad reflection characteristics. Each pigment reflects some energy from wavelengths across the visible spectrum (400 to 700 millimicrons).





PIGMENTS IN AN IMAGINARY WORLD, whose available light is limited to a band of wavelengths extending only from about 570 millimicrons to 590 millimicrons, would have to be much more sharply selective. Upper curves show reflection curves of pigments which would give full color in such a world. Lower curves represent the same curves stretched out so that the 570-590 band covers the same width as the 400-700 band of the visible spectrum.

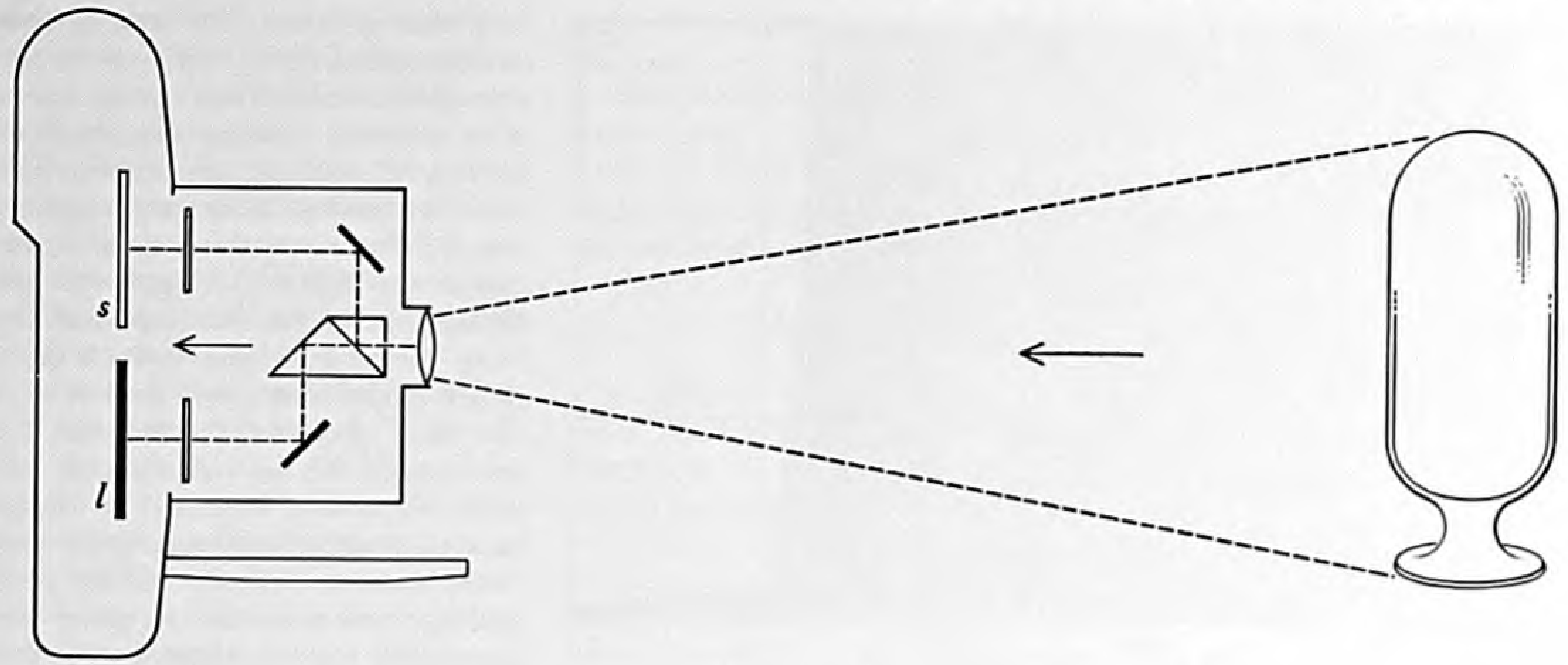
lengths [*see illustration at bottom of preceding page*]. We have also investigated the limits on relative brightness. With some pairs the colors are maintained over enormous ranges of brightness; with others they begin to break down with smaller changes. Again, the result depends on the wavelengths we are using. A table showing the stability of various colors for a sample pair of wavelengths appears on page 191.

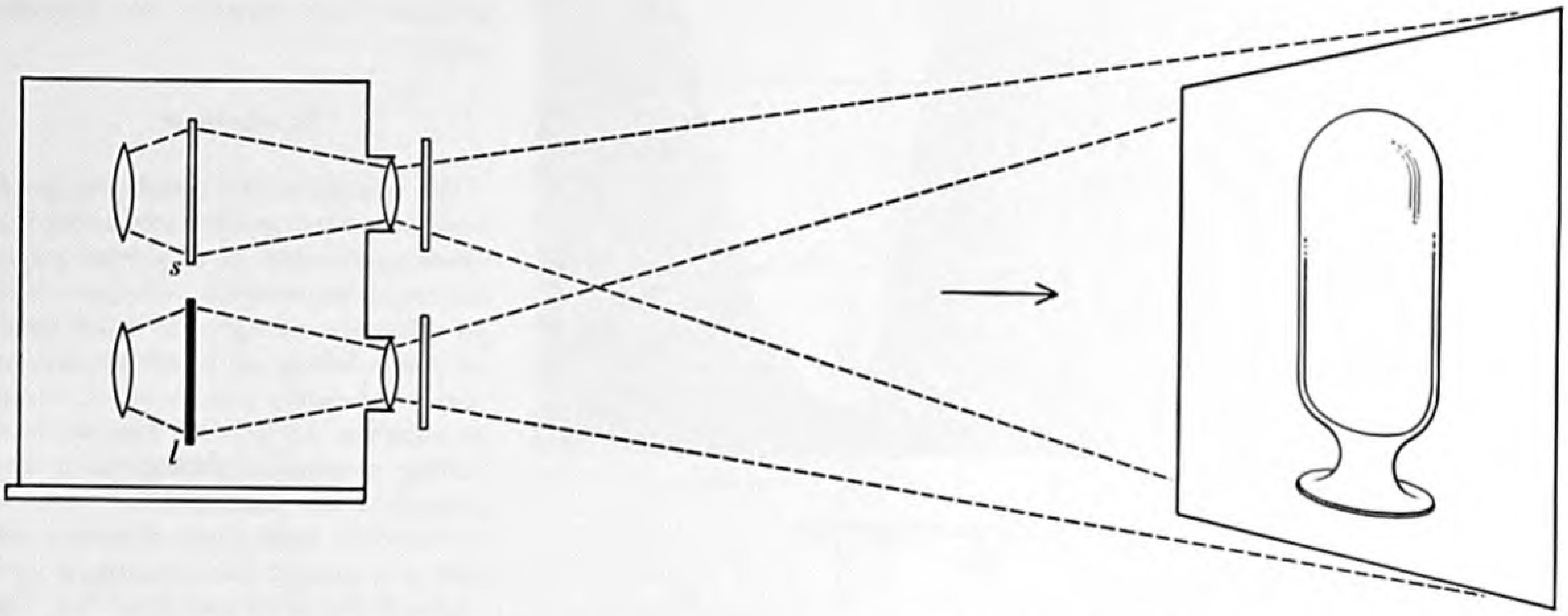
A New Coordinate System

The color map tells us what we will *not* see when we combine a pair of images at various wavelengths. Can we now make a positive prediction? Given a pair of records of the same scene, and a pair of wave bands with which to illuminate them, what color will appear at each specific point on the combined image? In other words, we want a set of rules that will do for images what the color triangle does for color-matching experiments (and what most of us have mistakenly supposed it does for images as well).

We have formed a new coordinate system that does for the first time predict the colors that will be seen in natural images. Perhaps the best way to approach it is through an actual experiment. Let us set up the dual projector (or the monochromator) for any pair of "long" and "short" bands, say red and white, that can produce full color. We know that local variations in the relative brightness of the two records must somehow give rise to the color. Yet we have also found that changing all the brightness ratios in a systematic way, for example by cutting down the total light from the red projector, has no effect. Therefore we look for a way of describing the brightness in terms that are independent of the total light available in either image.

This can be done as follows: We turn on the "long" projector alone, setting its brightness at any level. Now we find the spot on the red image corresponding to the point at which the long black-and-white record lets through the most light. We measure the intensity at that point and call it 100 per cent. It tells us the maximum available energy for the long waves. Next we measure the intensity of the light all over the rest of the red image, marking down for each point the red intensity as a per cent of the maximum available. Then we turn off the "long" projector, turn on the "short" one and follow the same procedure for the short wavelengths (in this case the full





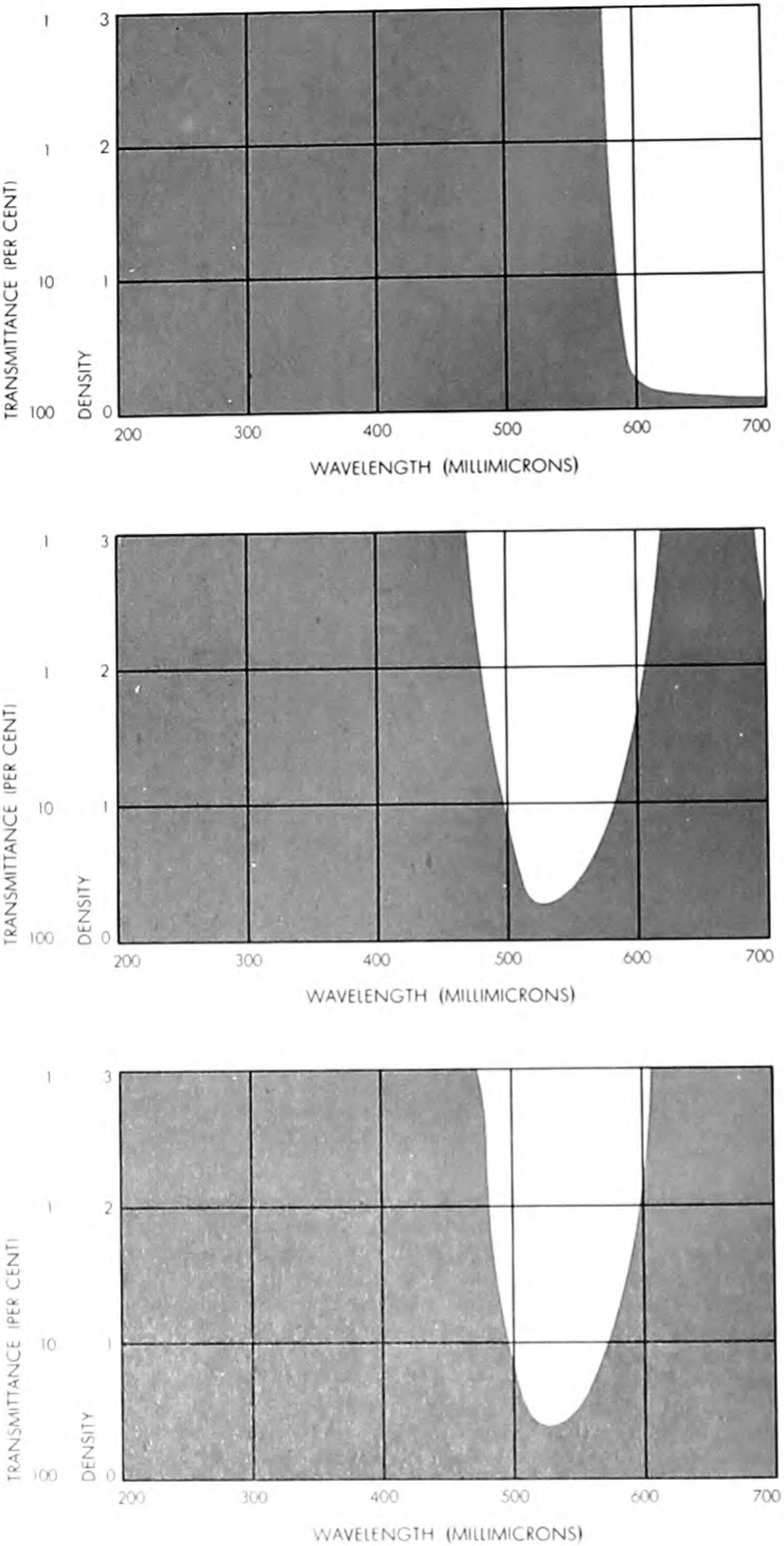
LONG AND SHORT RECORDS are prepared by photographing a scene with the dual camera diagrammed at top. Small open rectangles represent colored filters; the filter in front of the long record is red and the one in front of the short record is green. A composite image is formed by superimposing long and short records (*labeled l and s*) on a screen by means of a dual projector (*bottom*).

spectral band). Now we draw up a two-dimensional graph [*top of page 185*], plotting the percentage of available long wavelengths on one axis and the percentage of available short wavelengths on the other. Every point on the image can be located somewhere on this graph. Each time we plot a point, we note next to it the color it had on the image.

What emerges is a map of points, each associated with a color. When it is finished, we can see that the map is divided into two sections by the 45-degree line running from lower left to upper right. This is the line of gray points. If we had put the same transparency in each projector, all the points would fall on the gray line, since the percentage of available light would be the same at every point on the image for both projectors. The other colors arrange themselves in a systematic way about the 45-degree line. Warm colors are above it; cool colors are below. Thus it seems that the important visual scale is not the Newtonian spectrum. For all its beauty the spectrum is simply the accidental consequence of arranging stimuli in order of wavelength. The significant scale for images runs from warm colors through neutral colors to cool colors.

Repeating our experiment with different illuminating wavelengths or bands, we find that for every pair that produces full color the position of the colors on the coordinate graph remains the same. Thus we have the rule we were looking for, a rule that tells us in advance what color we shall find at any point in an image. We can take any pair of transparencies and measure their percentage of transmission in various regions of the picture. Then, before projecting them, we can predict the colors these areas will have. We will be right provided that the illuminating wavelengths are capable of stimulating all the colors. In cases where they are not, we must change the coordinate map accordingly. Thus the full set of rules consists of a group of coordinate color plots, one for each section of the color map at the bottom of page 185.

Note that each coordinate system is



WAVELENGTHS PASSED BY FILTERS used in various experiments described by the author are shown in these curves. At top is the transmittance curve for the red filter used in photographing long record; below it is transmittance curve of green filter for preparing short record. At bottom is curve for the green filter used in the sodium-viewer experiment.

itself dimensionless. The axes do not measure wavelength, brightness or any other physical unit. They express a ratio of intensities at a single wavelength or for a broad band of wavelengths. The axes have another interesting property: they are stretchable. Suppose we superimpose two identical long-wavelength photographs in the slide holder of the "long" projector and leave a single short-wavelength photograph in the holder of the "short" projector. We find that this combination still does not alter the colors on the screen. What sort of change have we made? Every point in the long record that transmitted 1/2 of the available light now transmits 1/4, points that transmitted 1/5 now transmit 1/25 and so on. On the logarithmic scale of our graph this corresponds to stretching the long-record axis to twice its former length. The 45-degree line now shifts to a new direction, but all the color points shift with it, maintaining their relative positions [*see diagram on opposite page*].

Randomness

Our studies of the coordinate graph have uncovered another interesting and subtle relationship. As we plotted graphs for various experiments we began to suspect that any arrangement which yielded points falling on a straight line, or even on a simple smooth curve, would be colorless. To test this idea we tried putting a negative photograph in one projector and a positive of that negative in the other. Such a pair of images will plot as a straight line running at right angles to the 45-degree gray line. The image is indeed virtually colorless, showing only the two "colors" of the stimuli involved in projection and a trace of their Newtonian mixture.

If an image is to be fully colored, its coordinate graph must contain points distributed two-dimensionally over a considerable area. But even this is not enough. The points must fall on the graph in a somewhat random manner, as they do in the plot of any natural scene. This requirement can be demonstrated in a very striking experiment. Suppose we put a "wedge" filter in the slide holder of the red projector. The effect of the filter is to change the intensity of the beam continuously from left to right. That is, when the red projector is on and the white projector off, the left side of the screen is red and the right side is dark, with gradations in between. Now we place a similar wedge, but vertically, in the white projector so

that the top of the screen is white and the bottom is dark. With both projectors turned on we now have an infinite variety of red-to-white ratios on the screen, duplicating all those that could possibly occur in a colored image. However, they are arranged in a strictly ordered progression. There is no randomness. And on the screen there is no color—only a graded pink wash.

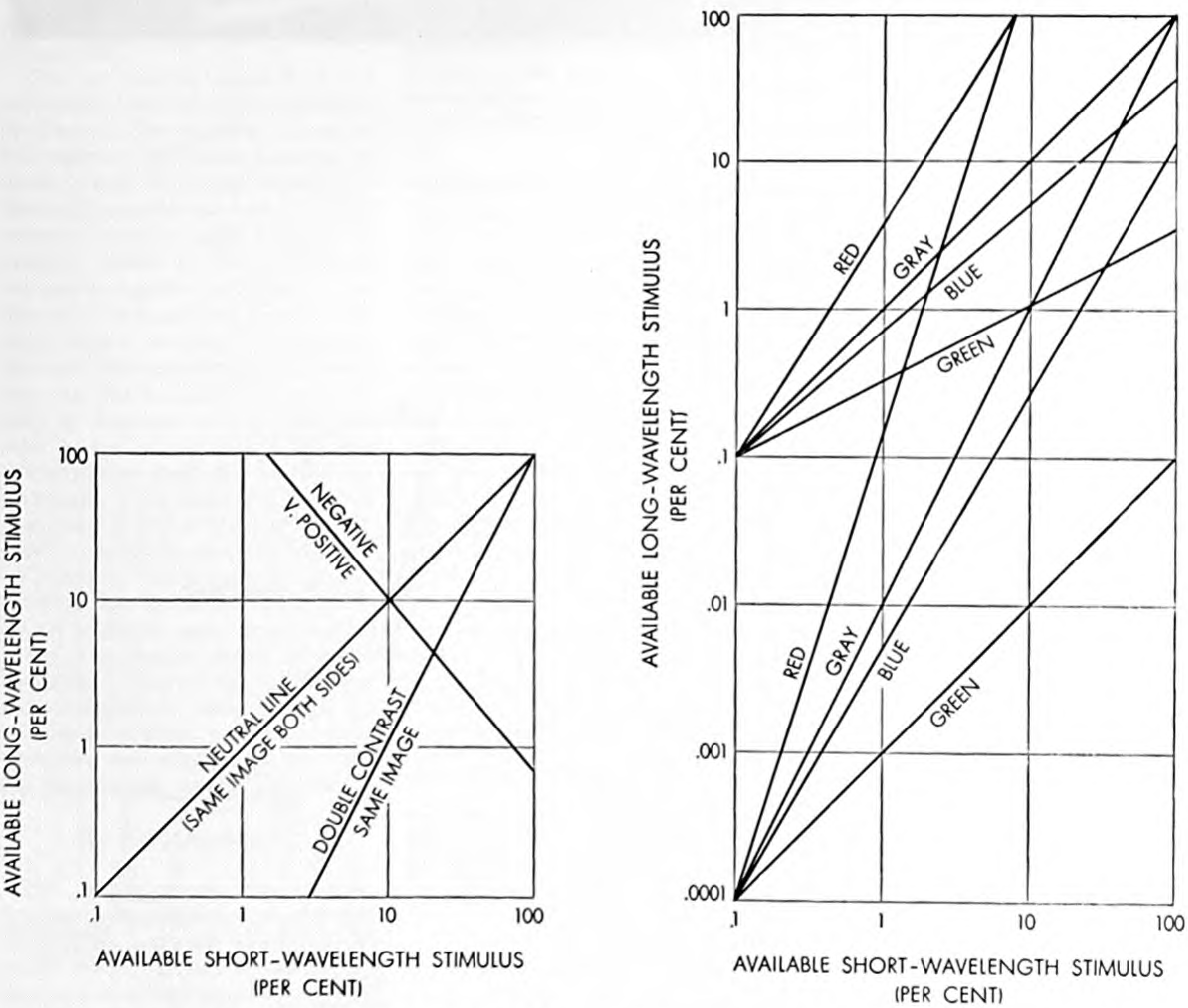
To repeat, then, the colors in a natural image are determined by the relative balance of long and short wavelengths over the entire scene, assuming that the relationship changes in a somewhat random way from point to point. Within broad limits, the actual values of the wavelengths make no difference, nor does the over-all available brightness of each.

The independence of wavelength and color suggests that the eye is an amazingly versatile instrument. Not only is it adapted to see color in the world of light in which it has actually evolved, but also it can respond with a full range of sensation in much more limited worlds. A dramatic proof of this is provided by another series of experiments.

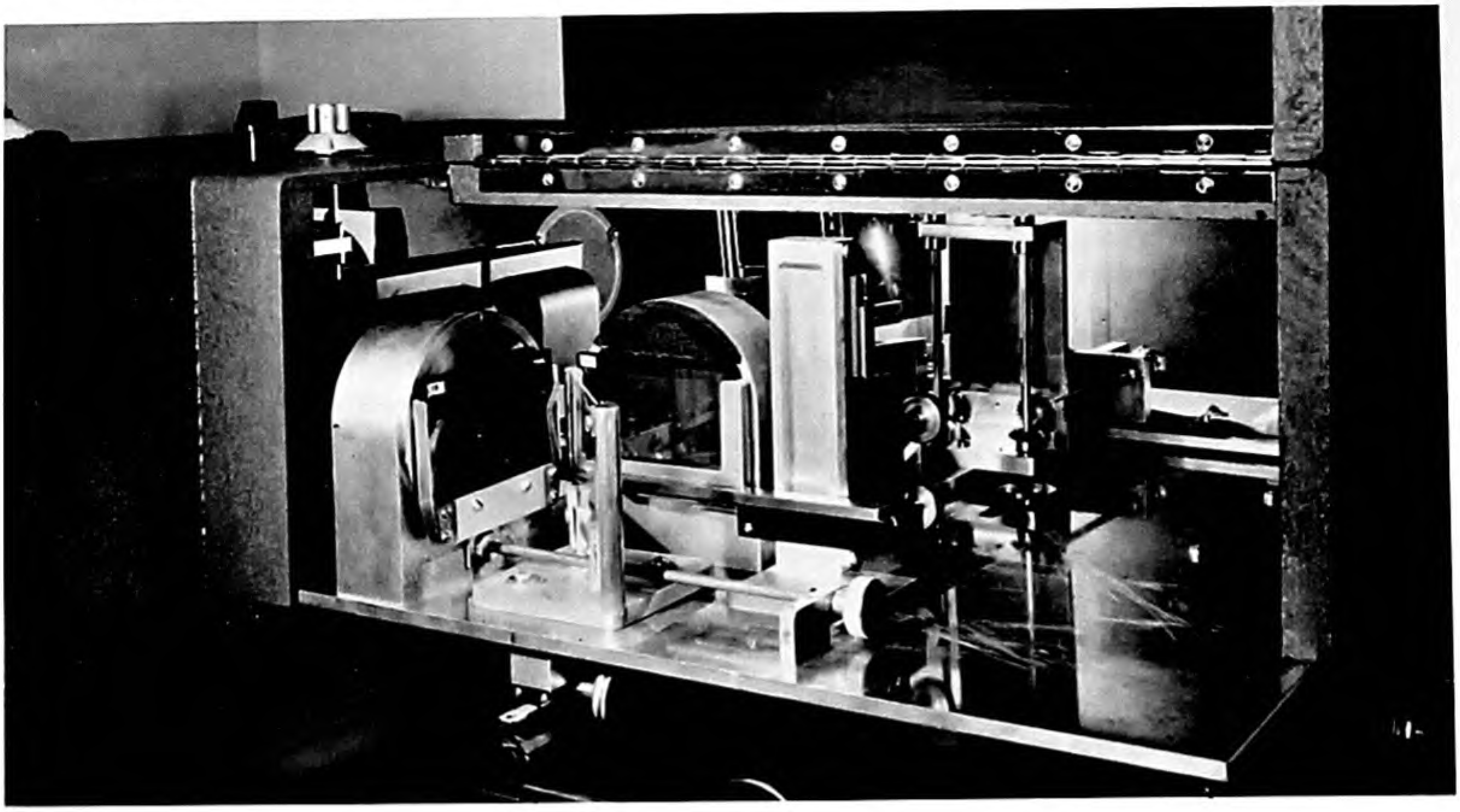
Color Worlds

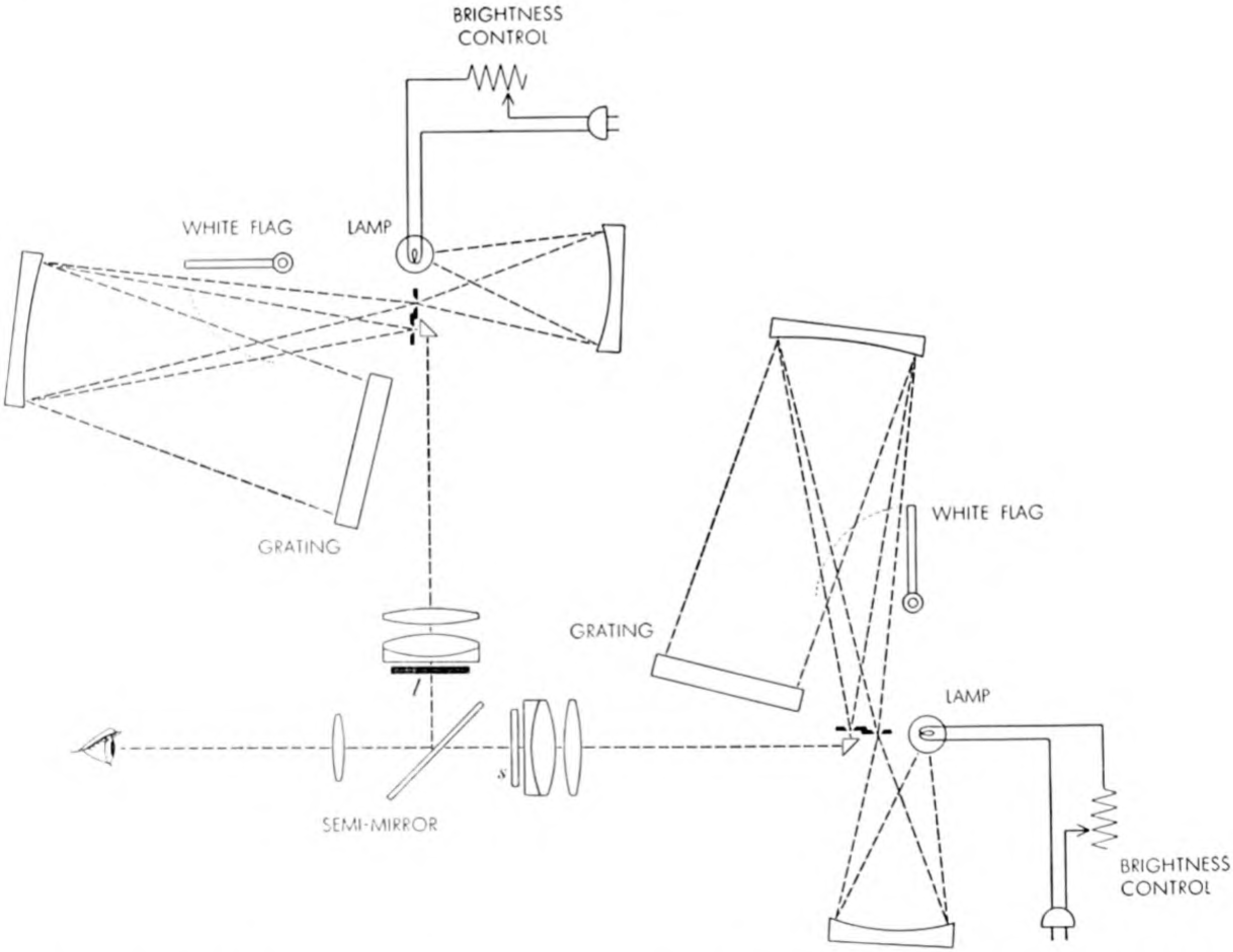
In these we use a pair of viewing boxes that superimpose fairly large images by means of semitransparent mirrors [*see diagram at bottom of page 191*]. Each box contains tungsten lamps, which produce white light, to illuminate one record and a sodium lamp to illuminate the other. We turn on one viewer, inserting the long and short transparencies and placing a red filter over the tungsten lamp. The composite image is fully colored, containing greens and blues, although the shortest wavelength coming from the mirror lies in the yellow part of the spectrum. Now we turn on the second viewer, inserting a green filter over the white light-source. Again the image contains a gamut of color, including red. The observer can see the images in both viewers at once—each showing the same range of color, but representing different visual worlds. In the first the sodium light (with a wavelength of 589 millimicrons) serves as the shortest available wavelength and helps to stimulate the green and blue. In the second it is the longest wavelength and stimulates red. If the observer stands back far enough from the viewer, he can also see the "natural" colors in the room around him. Here then is a third world in which yellow is "really" yellow.

Another way to use the green filter in the second sodium viewer is to hold it



PROPERTIES OF COORDINATE SYSTEM are illustrated in these diagrams. At left are the graphs of experimental situations which do not produce a gamut of color in the image. Such situations appear to graph as straight lines. At right the axes are shown to be stretchable. When gray line dividing warm and cool colors is displaced, colors move but maintain their relative positions.





DUAL MONOCHROMATOR is seen in schematic diagram at bottom and in photograph at top. Very narrow bands of wavelengths from any part of the visible spectrum are produced by the two gratings. White flags can be inserted to give white light. Narrow rectangles marked *l* and *s* represent the black-and-white transparencies that serve as the long record and the short record respectively.

up to the eye instead of placing it in front of the tungsten lamps. This filter passes both the sodium wavelength and the green band [*see bottom graph on page 188*]. When he looks around the room, the observer sees red objects as black and the rest of the colors as washed-out green. But when he looks at the picture in the second viewing box, he sees it quite full of color, including red.

The color worlds of the viewers are produced by pictures. Could we make physical models of these worlds, populating them with real objects which would show the same colors as the images in the viewers under the same conditions of illumination? We could if only we had the proper pigments. The pigments in the world around us are the best we have been able to find that look colored in our lighting: a spectrum of visible wavelengths from 400 to 700 millimicrons. Each of these pigments reflects a broad band of wavelengths, and its peak is not sharp [*see diagram at top of page 186*].

Thus our coloring materials do not distinguish clearly between wavelengths that are fairly close together. If we could find pigments with much narrower response curves, we would suspect that these might provide full color in a more restricted world of light—a world, for example, lighted by the wavelengths that pass through the green filter. In the absence of such coloring materials, we might content ourselves with creating this world photographically, if we could show that this is possible. A moment's study of diagrams on page 186 will show the exciting fact that a two-color separation photograph in a world of any band-width is the same as a two-color photograph in a world of any other band-width—including our own, provided that we postulate that a correctly proportioned change in the absorption bands of the pigments goes along with a change in the band-width of the world. Therefore we can use our regular long and short pictures, taken through the red and green filters, to transport ourselves into new worlds with their new and appropriately narrow pigments.

The Visual Mechanism

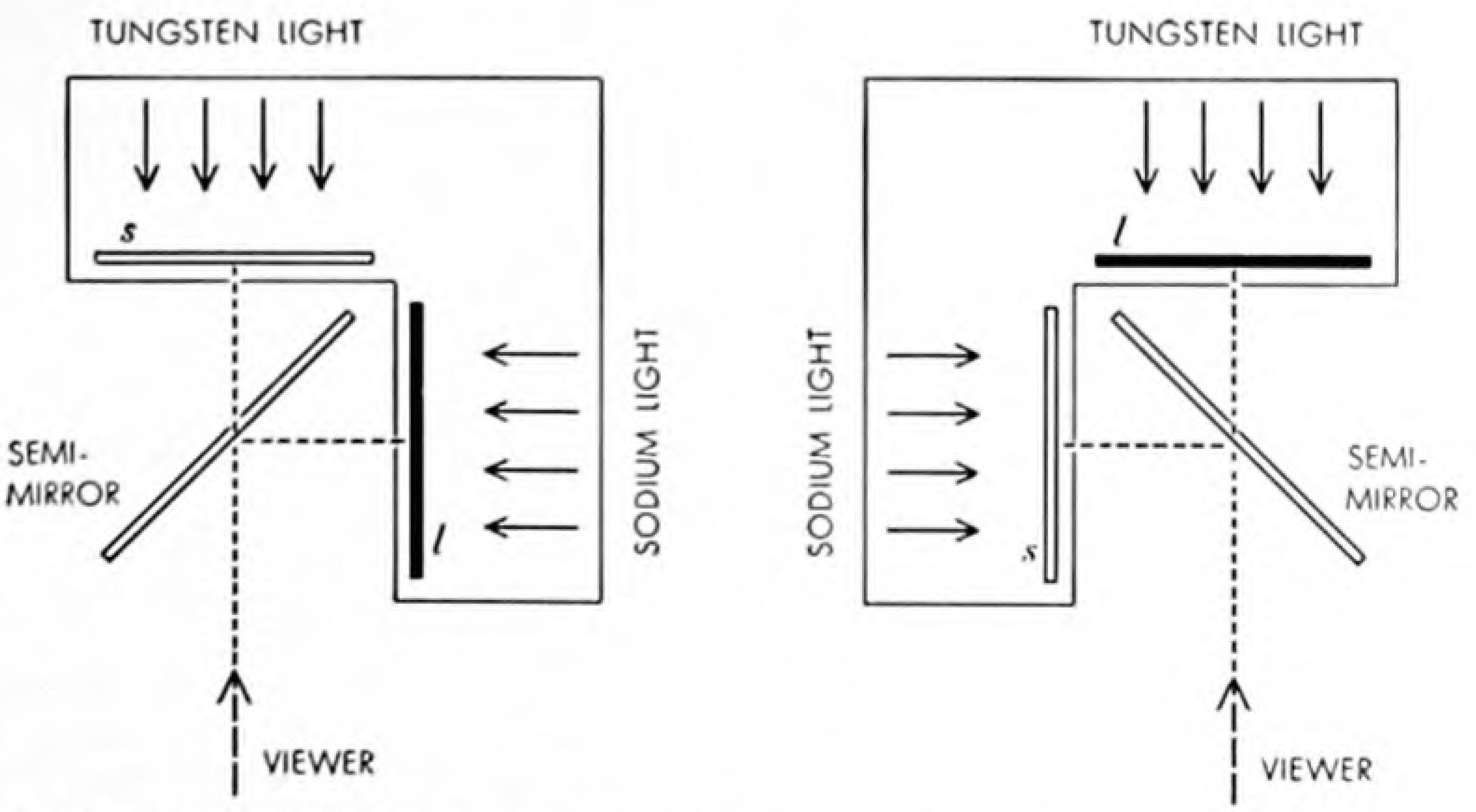
The sodium-viewer demonstration suggests an important consideration that we have not previously mentioned, although it is implicit in what has already been said. If the eye perceives color by comparing longer and shorter wavelengths, it must establish a balance point or fulcrum somewhere in between, so that all wavelengths on one side of it are taken as long and all on the other side as short. From the evidence of the viewer we can see that the fulcrum must shift, making sodium light long in one case and short in the other.

Where is the fulcrum in the ordinary, sunlit world? Experiments on a large number of subjects indicate that it is at a wavelength of 588 millimicrons. When we use this wavelength in one part of the dual monochromator and white light in the other, the image is nearly colorless. With a wavelength shorter than 588 millimicrons, white serves as the longer stimulus in producing color; with a wavelength longer than 588 millimicrons, white becomes the short record.

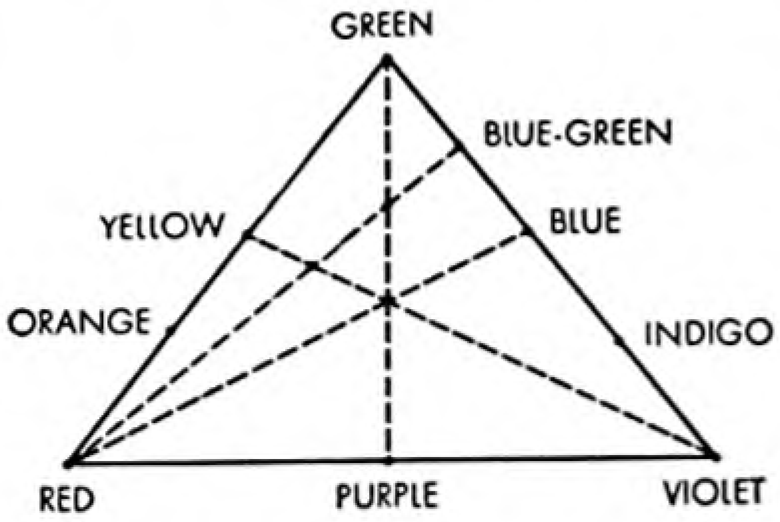
From the dual-image experiments we learn that what the eye needs to see color is information about the long and short wavelengths in the scene it is viewing. It makes little difference on what particular bands the messages come in. The situation is somewhat similar to that in broadcasting: The same information can be conveyed by any of a number of different stations, using different carrier frequencies. But a radio must be tuned to the right frequency. Our eyes are always ready to receive at any frequency

COLORS SEEN	RANGE OVER WHICH SEEN	VARIATION IN COLOR OVER THIS RANGE
GRAY	200 TO 1	LITTLE VARIATION
BROWN	100 TO 1	YELLOW-BROWN TO DARK BROWN
WHITE	100 TO 1	YELLOWISH-WHITE TO BLUISH-WHITE
YELLOW	30 TO 1	YELLOW TO OFF-WHITE
YELLOW-GREEN	30 TO 1	YELLOW-GREEN TO YELLOW ORANGE
BLUE	10 TO 1	BLUE-VIOLET TO BLUE GREEN
GREEN	6 TO 1	BLUE-GREEN TO GRAY-GREEN
RED	5 TO 1	DARK RED TO DARK ORANGE-RED
ORANGE	5 TO 1	YELLOW TO RED-ORANGE

LIMITS OF STABILITY of colors under variation in relative brightness of a sample pair of long and short stimuli are summarized in this typical chart. Second column shows the mechanism ratio (changing the brightness of either or both of the stimuli) for which color at left is recognizable. Pair of stimuli used was 450 millimicrons and 575 millimicrons.



SODIUM-VIEWING BOXES are diagrammed schematically. Each of these instruments produces a large composite image by means of the semitransparent mirror. Tungsten light is white, and is restricted to narrower bands of wavelengths by means of colored filters.



COLOR TRIANGLE of classical theory is shown in an early schematic form. Points of intersection of lines represent colors obtained by mixing spectral wavelengths in amounts proportional to distances from sides of triangle. Central point is equal mixture of primaries and is therefore white.

in the visible spectrum. And they have the miraculous ability to distinguish the longer record from the shorter, whatever the frequencies and the band-widths. Somehow they establish a fulcrum and divide the incoming carrier waves into longs and shorts around that point.

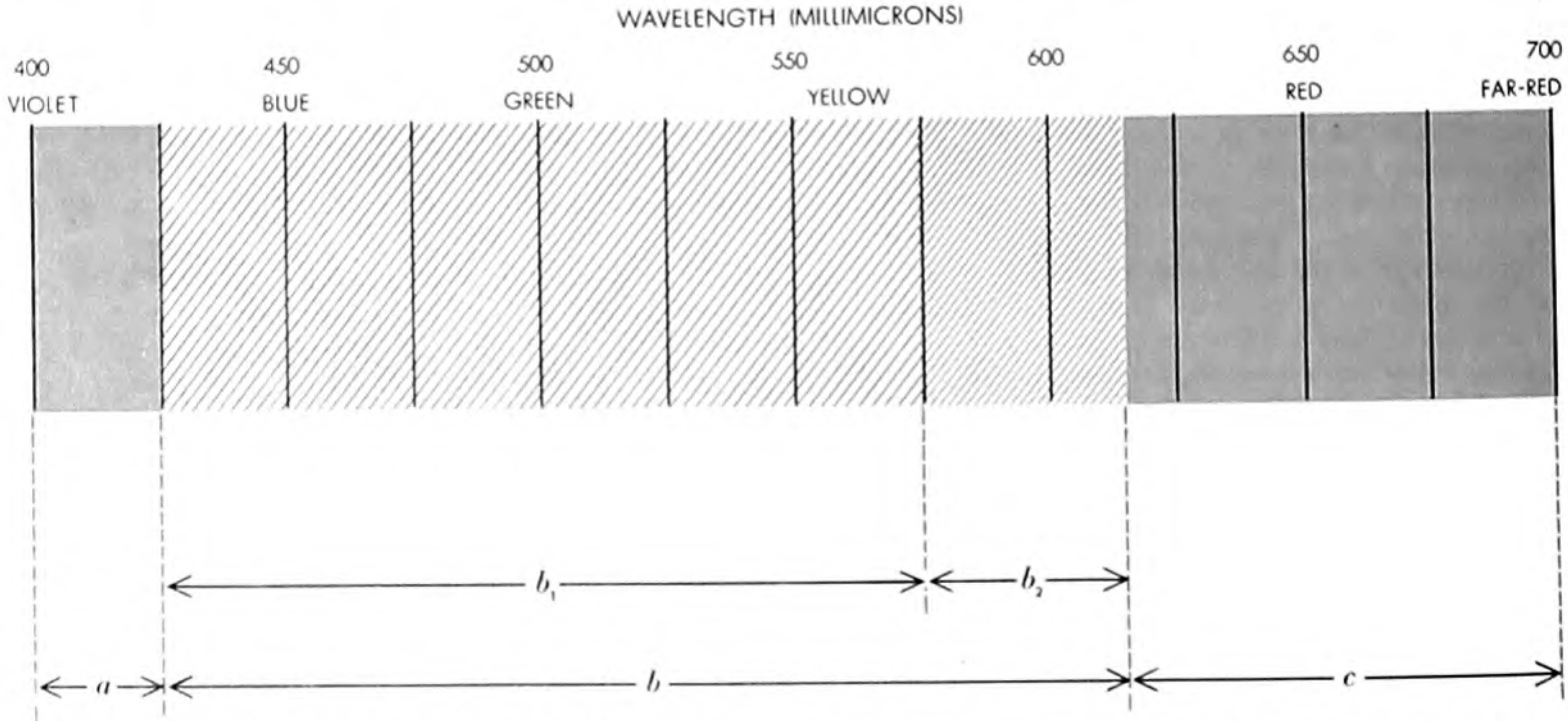
In our experiments we provide a single photograph averaging all the long wavelengths and a single photograph averaging all the short. What happens in the real world, where the eyes receive a continuous band of wavelengths? We are speculating about the possibility that these wavelengths register on the retina as a large number of individual color-separation "photographs," far more than the three that Maxwell thought necessary and far more than the two that we have shown can do so well. The eye-brain computer establishes a fulcrum wavelength; then it averages together all the photographs on the long side of the fulcrum and all those on the short side. The two averaged pictures are compared, as real photographic images are compared in accordance with our coordinate system.

Finally I should like to make clear that, although our experiments deal with two photographs and our coordinate system is two-dimensional, we have not been describing a two-color theory of vision. When we use a band of wavelengths for either or both of the records, we have light of many wavelengths coming from each point on the screen. And if classical three-color theory holds, it should describe the color of each of these points. This, as we have seen, it completely fails to do. It is true, however, that our experiments deal with two packages of information. We have demonstrated that the eye can do almost everything it needs to do with these two packages. The significance of what a third package will add is far from obvious. We are building a triple image-illuminating monochromator to find out.

A third picture may provide better information at the photographic level or an additional and useful interaction with the stimuli from two images. However, there is not a very big gap in the sensation scale to be filled by the third picture. In a given image a particular combination of two stimuli might not provide an electrically intense blue or a delicately yellowish green, but it is still likely to provide more than enough for the animal to live with. Nevertheless we do expect that the richness of many colors will be increased by the interplay of a third stimulus. Whatever we learn by adding a third picture, the visual process will remain an amazing one from the evolutionary point of view. Why has a system that can work so well with two packages of information evolved to work better with three? And who knows whether it will not work better still with four, or five or more?

What does the eye itself do in the everyday world of the full spectrum? Does it make only two averages? Or does it put to better use the new ability we have discovered—the ability to distinguish sharply between images at closely spaced wavelengths? Perhaps it creates many sets of averages instead of just two or three.

Even if more than two information channels are used, we feel that the big jump is obviously from one to two. Most of the capability of our eyes comes into play here. And whatever may be added by more channels, the basic concept will remain. Color in the natural image depends on the random interplay of longer and shorter wavelengths over the total visual field.



WAVELENGTH AND COLOR are independent of each other, except for the long-short relationship. This diagram shows the roles that various wavelengths can play. Those in the interval a can serve only as the short-record stimulus; those in b may be either long or short; those in c can only be long. If the wavelengths in b_1 are used as short-record stimuli, they will combine with a longer wavelength to produce the full gamut of color. If they are used as long-record stimuli, they will produce a more limited range. Wavelengths in b_2 will produce full color, serving as the stimuli for either the long record or for the short record. When both stimuli come from between 405 and 520 milli-microns, "short-wave reversal" occurs (*see color map on page 192*).

The Author

EDWIN H. LAND is president, chairman of the board and director of research of the Polaroid Corporation, a manufacturing concern that he founded in 1937. To the general public Land's name is best known in association with the popular camera that presents the photographer with a finished positive print within seconds after the shutter is clicked. Before he developed the Land camera he had devoted most of his time to "Polaroid" polarizing filters, on which he had begun to work while he was still an undergraduate at Harvard University.

Bibliography

COLOR VISION AND THE NATURAL IMAGE. PART I. Edwin H. Land in *Proceedings of the National Academy of Sciences*, Vol. 45, No. 1, pages 115–129; January, 1959.

COLOR VISION AND THE NATURAL IMAGE. PART II. Edwin H. Land in *Proceedings of the National Academy of Sciences*, Vol. 45, No. 4, pages 636–645; April, 1959.

SYMPOSIUM ON NEW DEVELOPMENTS IN THE STUDY OF COLOR VISION. *Proceedings of the National Academy of Sciences*, Vol. 45, No. 1, pages 89–115; January, 1969.

SENSATION AND PERCEPTION IN THE HISTORY OF EXPERIMENTAL PSYCHOLOGY. Edwin G. Boring. Appleton-Century-Crofts, Inc., 1942.

SCIENTIFIC AMERICAN June 1958, Vol. 198, No. 6, pp. 30-36 OFFPRINT 224

SUPERFLUIDITY

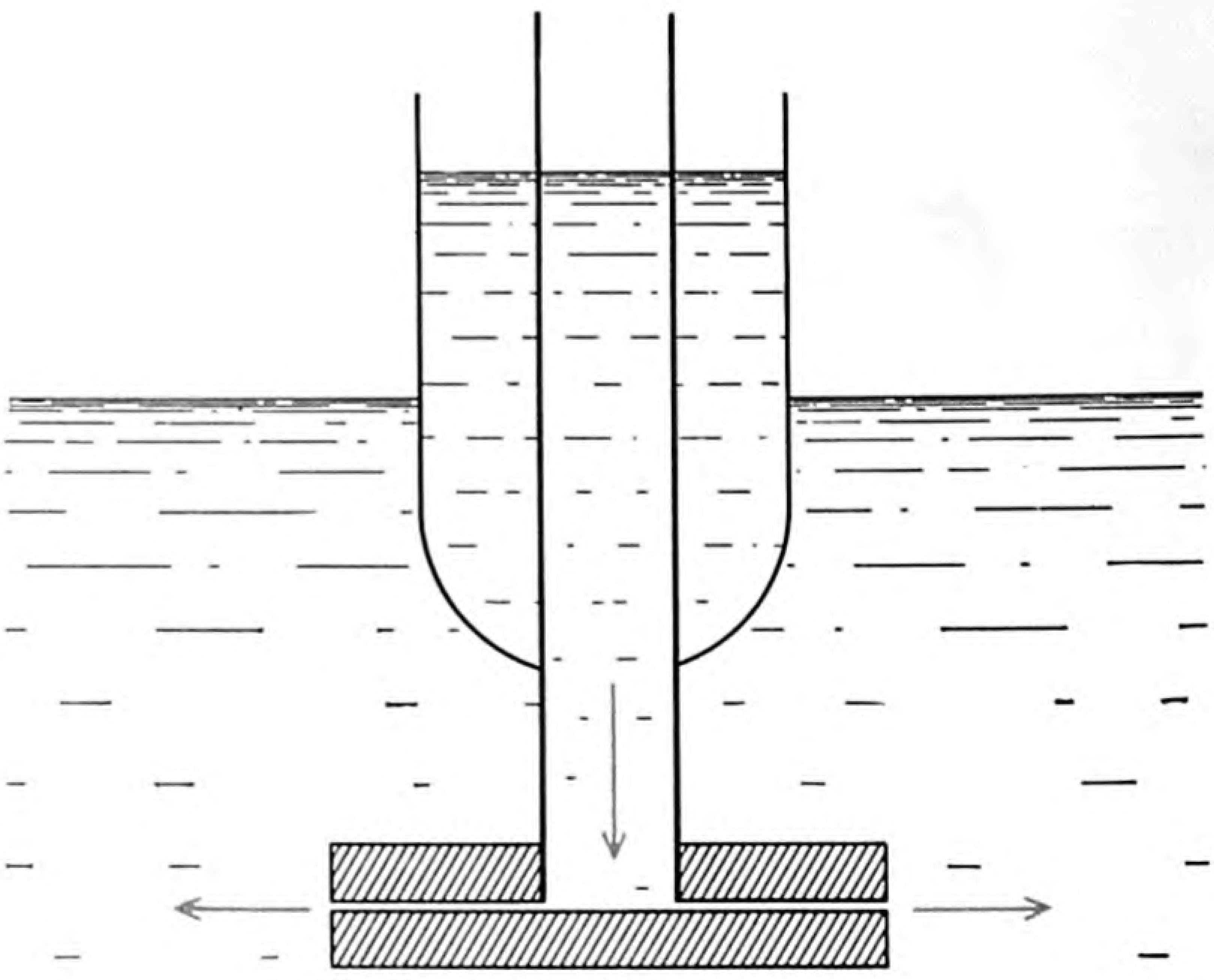
by Eugene M. Lifshitz

When liquid helium is cooled to 2.2 degrees above absolute zero, it flows without friction. A Soviet physicist describes how this strange fluid exhibits the quantum properties of individual atoms.

The history of physics is marked by breakthroughs into new, unexpected worlds of discovery. One such event occurred just half a century ago, on July 10, 1908, the day Heike Kamerlingh Onnes of the University of Leiden succeeded in cooling helium gas down to the liquid state. His breakthrough into this low temperature region, close to absolute zero, led to the discovery of two strange properties of matter, totally different from those familiar to us at ordinary temperatures. Both of these properties involve the phenomenon of "frictionless" flow. One of them, soon discovered by Kamerlingh Onnes himself, is superconductivity, the frictionless flow of electrons and the consequent complete disappearance of a metal's resistance to electric current [see "Superconductivity," by B. T. Matthias; SCIENTIFIC AMERICAN Offprint 227]. The other, not discovered until 30 years later, is the superfluidity of liquid helium, the frictionless flow of entire atoms, demonstrated in the liquid's ability to flow through the tiniest tubes or narrowest slits. Our understanding of this remarkable phenomenon is today almost complete. I have been asked by the editors of SCIENTIFIC AMERICAN to give a short survey of what has been learned about superfluidity, first discovered in 1937 by Peter L. Kapitza at the Institute for Physical Problems in Moscow.

Helium is the only substance which remains liquid under ordinary pressure at absolute zero; all others freeze to the solid state if sufficiently cooled. Now according to ordinary or "classical" concepts, at absolute zero all the atoms of a substance cease moving and occupy fixed positions within the body, thus causing it to become solid. The fact that helium remains liquid is the first indication that its properties can be understood only in terms of the entirely different concepts of quantum mechanics.

As students of physics know, quantum mechanics is the peculiar system of laws that governs the properties of matter on the microscopic scale--the world of individual atoms and molecules. But in the case of liquid helium we can see matter displaying "quantum properties" on the macroscopic scale—that is, in its bulk or gross form, where we deal with an enormous number of atoms. We must pause to consider what this means. All gross matter possesses not only the ordinary properties with which we are familiar but also fundamental quantum properties. At ordinary temperatures the random heat motion of atoms and molecules cloaks the quantum behavior of the individual particles, and matter exhibits only its familiar properties; that is, it obeys the rules of ordinary classical mechanics. The quantum theory says that if we reduce a substance to extremely low temperatures, so that its atoms' heat motions become very weak, the fundamental quantum properties of the substance should become observable. However, all substances except one solidify before the quantum properties



EXPERIMENT BY PETER KAPITZA demonstrated the superfluidity of helium II in its ability to flow through an extremely narrow slit between two highly polished glass plates. With the slit narrowed to half a micron, the flow of helium II quickly equalized the levels in tube and reservoir, while the flow of helium I through the slit was scarcely perceptible.

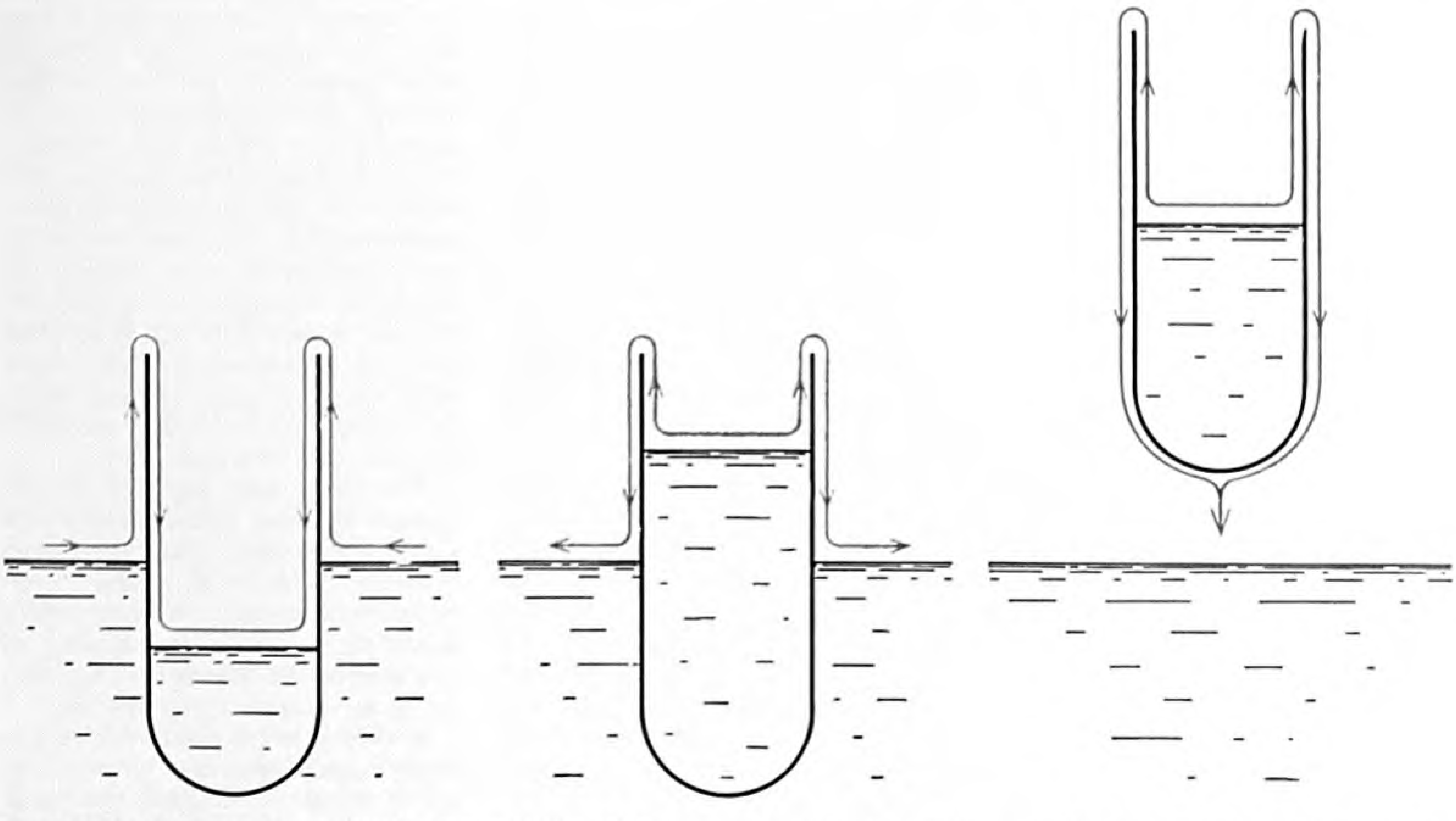
emerge. The single exception is helium: it succeeds in becoming a "quantum" substance before solidification. Once this has happened it need no longer solidify at all, since one of the principles of quantum mechanics is that the motion of atoms need not cease completely at absolute zero. Thus in liquid helium nature has made a "quantum liquid" available to physicists for investigation.

Helium liquefies at 4.2 degrees Kelvin (4.2 degrees centigrade above absolute zero). Upon being cooled to 2.2 degrees Kelvin, helium, although remaining liquid, undergoes another type of transition. It was first observed as an abrupt jump in the liquid's specific heat (*i.e.*, heat capacity). Because the graph showing this abrupt change has the shape of the inverted Greek letter lambda [*see chart at top of next page*], the point of transition (2.2 degrees) became known as the "lambda point." Liquid helium above this point was called helium I; liquid helium below it, helium II. It is the latter which was found to be a liquid with unique properties.

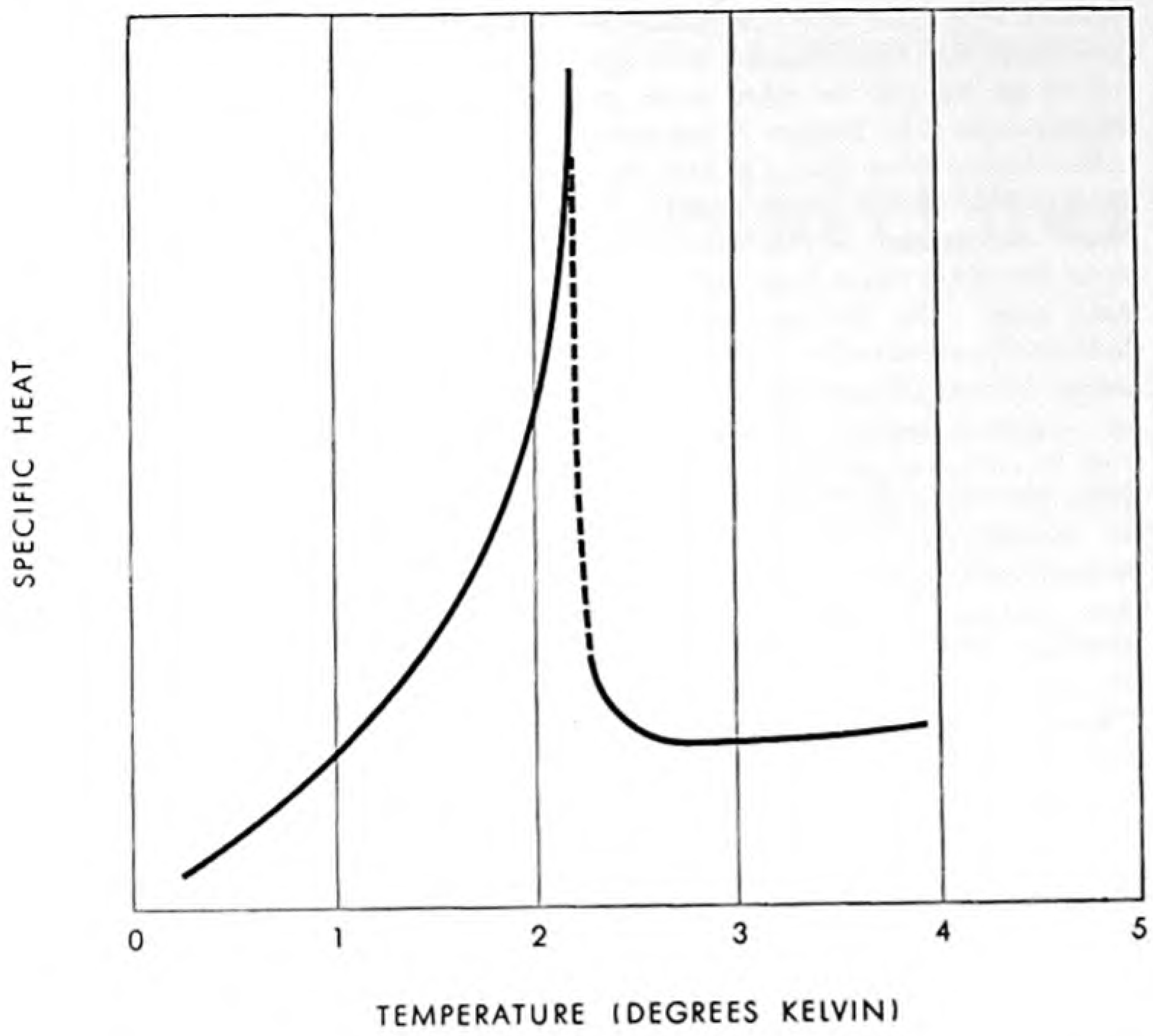
The first indication of these properties came in 1935 from the Kamerlingh Onnes laboratory. W. H. Keesom and his sister Miss A. P. Keesom discovered that helium II is an extraordinary conductor of heat. Experimenting with a capillary tube filled with the substance, they found that heat traveled from one end of the tube to the other at an extremely rapid rate. Helium II appeared to be a far better conductor of heat than the best of known conductors under ordinary conditions: it conducted heat about 200 times faster than copper normally does. (The Keesoms' discovery, incidentally, explained an odd feature of helium II's previously observed behavior: liquefied helium, absorbing heat from its container, bubbles like boiling water, but when the liquid is cooled to the lambda point, it suddenly becomes perfectly still. The reason now became plain: helium II conducts heat away from the walls of the container so rapidly that no bubbles form at the walls, as they do in ordinary boiling. The liquid vaporizes only at its open surface.)

It was an attempt to explain helium II's remarkable conduction of heat that led Kapitza to discovery of its superfluid property. Why did heat travel so rapidly through this material? Kapitza suspected that the swift transport of heat in helium II was due not to any exceptional conductivity of the substance but to movement of the liquid itself, in other words, to what are called convection currents. If this was so, helium II must be extraordinarily fluid; in the terms of physics, it must have an extremely low viscosity, meaning an extremely small internal frictional resistance to flow. The viscosity of a liquid is usually measured by letting it flow through a narrow capillary tube. In the present case, however, such measurement was found to be impracticable and it was necessary to devise a special apparatus involving the flow of a larger amount of liquid than is passed by a narrow capillary. Kapitza achieved success by letting the liquid flow between two polished glass disks forming a slit as narrow as half a micron (a fifty-thousandth of an inch). He found that whereas helium above the lambda point scarcely flowed through this slit at all, helium II passed through it very rapidly. In fact, he arrived at the amazing conclusion that the viscosity of helium II was less than a ten-thousandth that of hydrogen gas! On the basis of his measurements Kapitza daringly suggested that helium II had no viscosity at all, and exhibited a new phenomenon which he called "superfluidity."

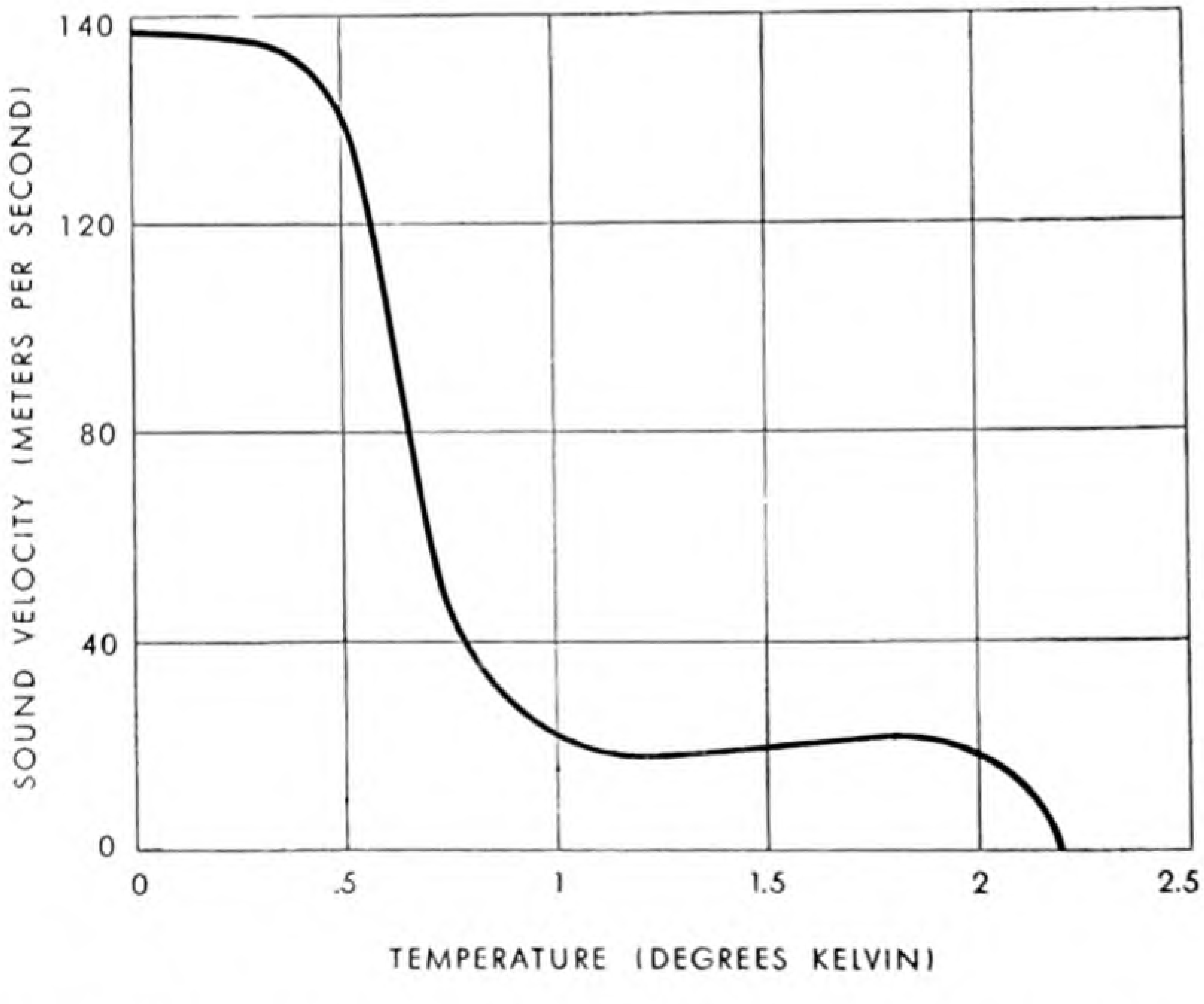
Directly related to the superfluidity of helium II is the remarkable phenomenon of the "creeping film." It had been noted long before that, if liquid helium is poured into a flask separated into two chambers by a partition, somehow the two levels spontaneously manage to equalize in course of time. John



CREEPING FILM provides spectacular demonstration of the superfluidity of helium II. Any liquid that wets a surface forms a film. In the case of helium II, however, the film forms quickly and behaves more or less as a siphon along which the liquid flows. As shown here, it will flow into a flask lowered into a helium reservoir (*left*) or out of it (*center*) or out of a flask withdrawn from the helium bath (*right*) to escape in the form of drops from its bottom. The velocity of this creeping film may be a foot or more per second.



LAMBDA POINT at 2.2 degrees Kelvin marks the transition of liquid helium from the fluid to the superfluid state. At this point there is a sharp drop in heat capacity of the liquid.



VELOCITY OF SECOND SOUND in helium II ascends at temperatures below 1 degree. The velocity of ordinary sound, at 240 meters per second, is hardly affected by temperature.

G. Daunt and Kurt Mendelssohn at the University of Oxford demonstrated by direct experiment that helium flows between the containers by means of a film, some millionths of an inch thick, formed on their walls [*see diagram on preceding page*]. The liquid in the film may travel at a speed of more than a foot per second. The propensity to form a film is in itself not a unique property of helium II. Such films are formed by any liquid which wets a solid surface, but the viscosity of an ordinary liquid is such that the film forms slowly and moves scarcely at all. Helium II is the only fluid which, owing to its superfluidity, forms a swiftly moving film.

Investigation soon disclosed other paradoxical features of helium II's behavior which were difficult to explain. For one thing, experimenters at Leiden and at the University of Toronto found that, although in its flow through a narrow slit it behaved as if its viscosity was zero, the liquid did show some viscosity, albeit small, in another experiment: when a cylinder or disk was rotated in the liquid, the fluid exerted a measurable frictional resistance to the rotation of the body.

Kapitza observed a still more puzzling phenomenon. It emerged in the course of some experiments on the transport of heat in helium II. Seeking to demonstrate that heat was transported via movement of the liquid, he used a flask with a movable vane suspended in front of its opening, so that any outflow of liquid would deflect the vane. He filled the flask with liquid helium and immersed it in a helium bath. When he heated the liquid in the flask (by shining light on a blackened absorbing surface in the flask), the vane was deflected, showing that a stream of liquid was flowing out. This gave a clear proof that the flow of heat in helium II involved some sort of movement of the liquid. But Kapitza was struck by a paradoxical fact. Although liquid flowed out of the flask, the flask remained full!

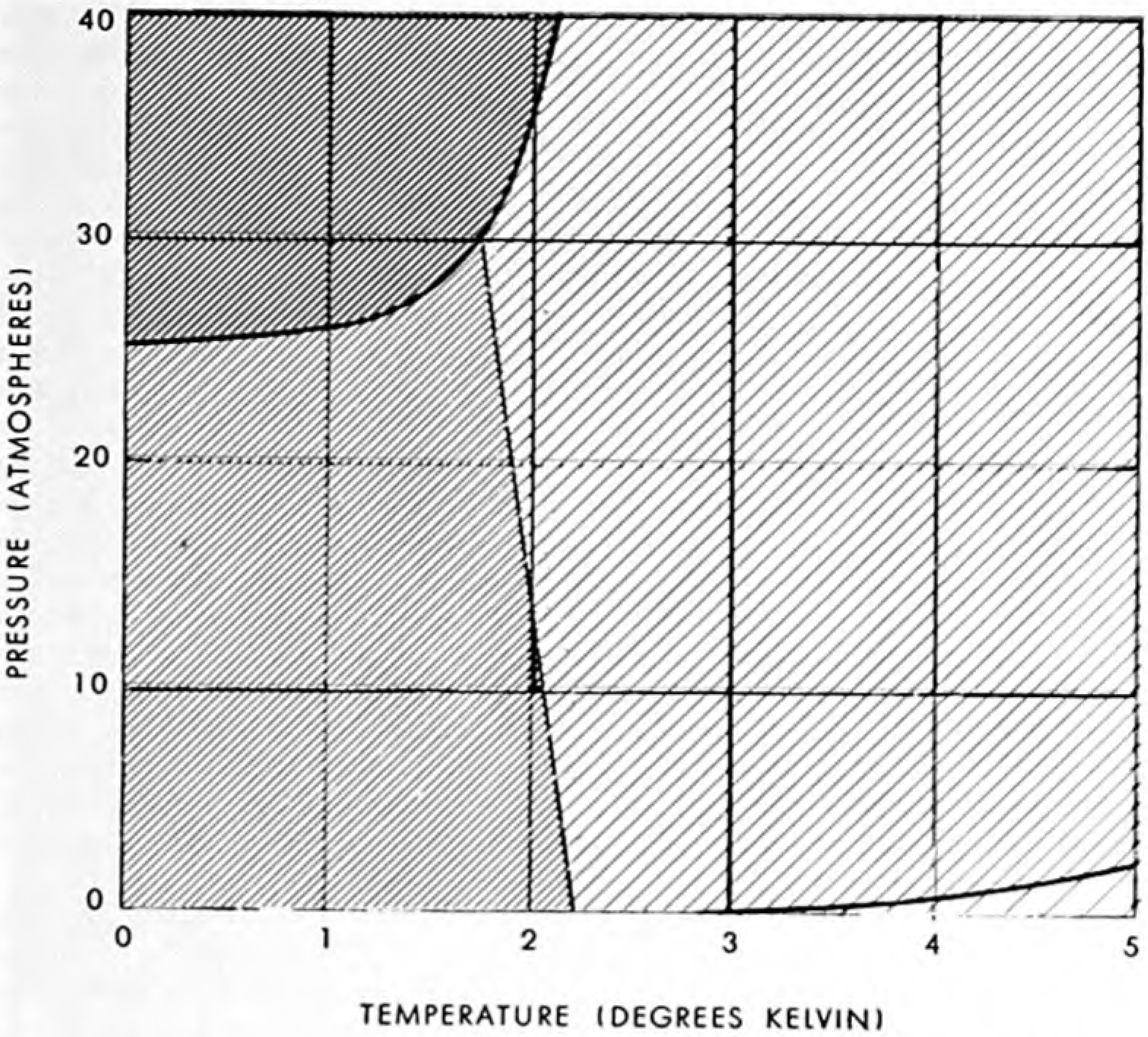
How could these paradoxes be explained? Here was a fluid which flowed as if it had no viscosity and yet showed a definite viscosity in certain experiments; it was a fluid, furthermore, which flowed out of a flask upon heating, as indicated by the deflection of the vane, yet its flow did not empty the flask!

In 1940-1941 an elaborate theory to explain superfluidity was advanced by Lev D. Landau at Kapitza's institute in Moscow. Some qualitative features of the theoretical picture were independently predicted by Laszlo Tisza at the Collège

de France in Paris (he is now at the Massachusetts Institute of Technology).

The theory is quantum-mechanical in its nature, and it rests upon concepts in quantum mechanics which are understandable only to a specialist in the subject. But we can describe the remarkable picture of what takes place in helium II in fairly simple physical terms. We must bear in mind, however, that description in terms of everyday experience cannot really picture the quantum world, which is beyond our direct observation.

Helium II, says Landau's theory, can undergo two motions simultaneously. This is, of course, a statement that defies common sense. In the case of an ordinary liquid, we are able to derive a complete description of its flow by measuring the distribution of velocities in one direction in a cross section of the stream. That is how engineers measure the flow of water in a canal. In the case of liquid helium, however, a complete description of its flow necessitates the knowledge of not one but two velocities at each point.



PHASE DIAGRAM of helium shows its transition from the gaseous *(below solid line at right)* to the fluid *(to right of broken line)*, the superfluid *(to left of broken line)* and the solid *(above solid line at upper left)* states in relation to temperature and pressure.

This situation can be visualized by an analogy called the "two-fluid model." It pictures helium II as composed of two liquids which can move independently "through" each other without any mutual drag. (In reality, of course, there is only one liquid; we must remember that the "two-fluid model" is an analogy.) The two types of motion in helium II have entirely different properties. One of the "components" moves as if it had no viscosity; Landau called this the "superfluid component." The other component moves like an ordinary viscous liquid and is called "normal."

But this does not exhaust the differences between the two kinds of motion in helium II. The most important one is that the normal component transports heat, whereas the superfluid motion is accompanied by no heat transfer whatsoever. One can say, in a sense, that the normal component is the heat itself. Thus the heat acquires a sort of independence in helium II; it separates from the mass of the liquid and acquires ability to move relative to a "background" which itself is at absolute zero. This is in radical contrast to the usual idea of heat as the chaotic motion of atoms, inseparable from the mass of a substance.

These concepts make it possible to explain the strange results of some of the experiments with helium II. To begin with, we can account for the paradox that liquid helium shows no viscosity in its flow through a narrow slit but does display frictional effects on a rotating disk. In the first experiment it is the superfluid component that flows freely through the slit, whereas the viscous normal component is held back and leaks through very slowly; this experiment demonstrates the absence of viscosity in the superfluid component. In the second experiment it is the normal component that is responsible for the friction on the disk; this experiment accordingly measures the viscosity of the normal component.

But we can now draw a further conclusion: Since superfluid flow carries no heat, flow through a slit may be said to filter out "heatless" liquid, leaving the heat in the container. In an ideally thin slit the outflowing fluid should be at absolute zero. In an actual experiment one may expect this fluid to have a lower temperature than that in the container, although not zero. An effect of this kind was observed as early as 1939 by Daunt and Mendelssohn. Kapitza himself was able to show that the fluid pressed out of a container through a fine filter had a temperature as much as three or four tenths of a degree lower than that left behind—a considerable drop for a liquid which had a temperature of only one or two degrees absolute to start.

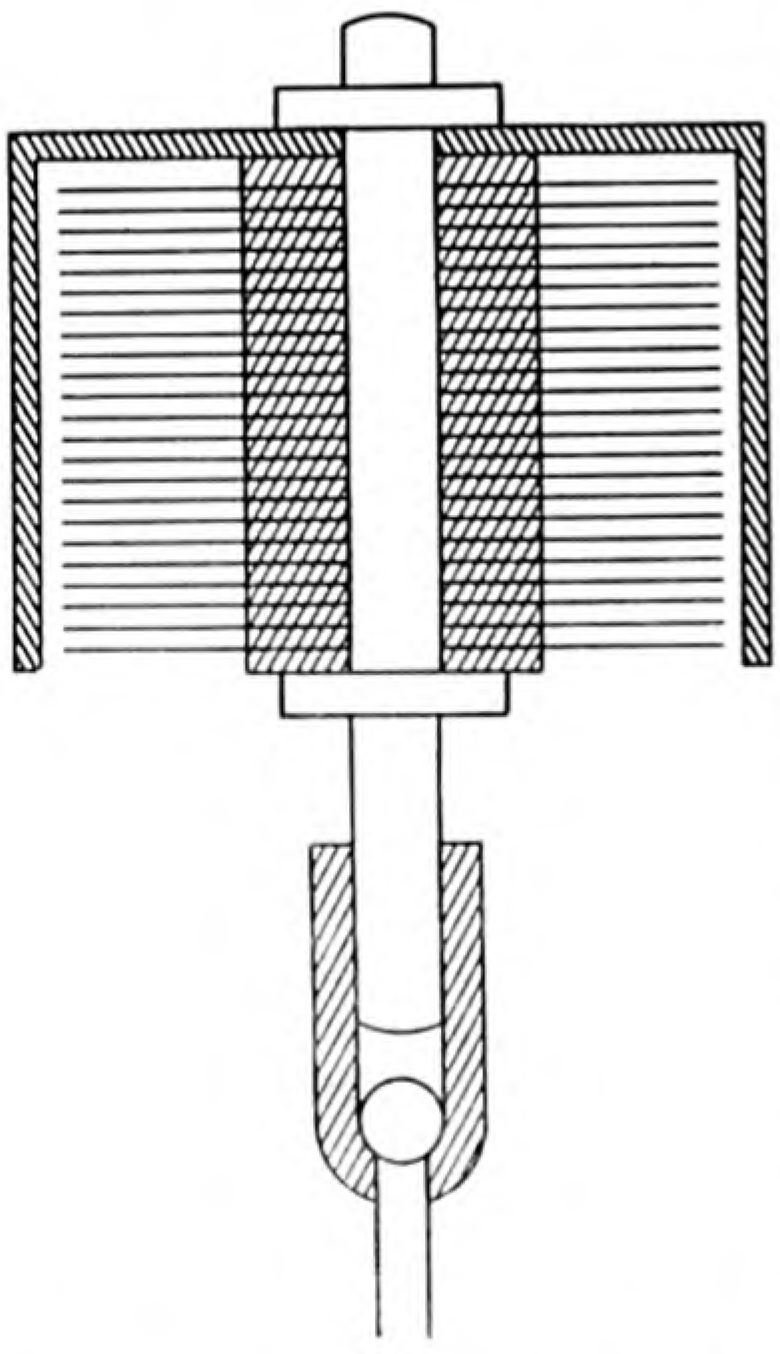
Landau's theory also explained, in equally simple fashion, the experiment in which Kapitza had caused helium II to flow out of a container and deflect a vane. The heat-bearing fluid that flowed past the vane, causing it to deflect, was the normal component; the flask remained full because a superfluid countercurrent flowed back in.

It remains to explain why this countercurrent did not, for its part, exert a force on the vane, thus counterbalancing the force exerted by the flow of the normal component. The reason is connected with another prediction of theory, the so-called "irrotational" character of the superfluid flow. The exact meaning of this word is not easy to explain in simple terms. What is remarkable is that as early as the 18th century the Swiss mathematician Leonhard Euler, who also worked in hydrodynamics, had predicted on theoretical grounds that the irrotational flow of an ideal nonviscous liquid past a solid body should exert no force on the body. Thus we arrive at a complete explanation of an unusual situation: With the normal component flowing in one direction and the superfluid component flowing in the opposite direction, one can say that there is no net motion of the liquid as a whole; yet the

vane is deflected, because one of the two components exerts no force.

An elegant experiment performed by Kapitza in 1940 confirms this explanation in a spectacular manner. He fashioned a tiny glass turbine with six bent capillaries radiating from it like the legs of a spider [*see photographs on opposite page*]. Immersed in a bath of helium II and heated by light, this turbine spun rapidly on its pivot, attaining speeds up to 120 revolutions per minute. This clearly evidenced the reactive force of the outflowing normal component. Kapitza now mounted a spider of silver wire on the turbine with vanes to oppose the outflow of helium from the mouth of each capillary. Because the reactive force of the normal component was offset by its push against the vanes, the spinning stopped. If the inflowing superfluid component could exert a force, the turbine would thereupon have begun to spin in the opposite direction. It remained motionless, in beautiful fulfillment of Euler's prediction and Landau's theory.

Landau's theory was verified in still another way. Since the superfluid component of helium II is capable only of irrotational flow, the theory predicted that the superfluid should not rotate as a whole in a whirling container (the way water in a glass rotates when the glass is twirled). E. L. Andronikashvili at Kapitza's institute put this prediction to the test in an experiment using rotating disks to drag the liquid [*see diagram on this page*]. The rotation of the liquid proved much less than would normally be expected, indicating that only the normal component of the helium rotated, while the superfluid component remained at rest. This experimental device, moreover, made it possible to measure the relative amounts of the normal and superfluid components in a given quantity of helium II. The ratio depends on the temperature of the liquid, as the theory predicts. Above the lambda point the liquid is entirely normal. At the lambda point the superfluid component begins to appear, and its "amount" increases as the temperature drops further. At absolute zero helium II should become entirely superfluid.



IRROTATIONAL FLOW of helium II was demonstrated by rotation of stack of thin foil disks in a bath of the superfluid. The drag on the disks was much lower than would have been expected in a normal fluid.

Finally let us consider the phenomenon called "second sound"—a property of helium II which was predicted independently both by Landau and Tisza. The prediction was that two different kinds of waves, propagated at different velocities, could travel through the liquid. According to Landau's theory the two kinds of wave propagation arise from the fact that helium II is capable of performing two motions simultaneously. If both components of the liquid oscillate jointly—that is, if the superfluid and normal components move in unison in the same direction—a sound wave of the ordinary kind arises, such as normally travels through a liquid. But, says the theory, there could also be a second kind of wave, specific for helium II, arising from oscillatory movement of the two components in opposite directions, each passing "through" the other in an oscillation cycle. It turns out that this wave should travel through the liquid at another (in fact, a much slower) velocity than ordinary sound, and that therefore it should be detectable as a second signal.

The phenomenon proved, however, to be difficult to discover experimentally. The first experiments, performed at our institute in Moscow in 1940, failed. Sound waves generated by vibration of a piezoelectric plate were passed through a tube filled with helium II. But only one signal arrived at the other end: no second signal could be detected.

Reviewing the problem in 1944, I found an explanation for the failure of the experiments. Ordinary sound waves, as is well known, are propagated in the form of cyclical compressions and rarefactions which move through the material medium (gas, liquid or solid). But analysis made clear that in the "second sound" wave, with the two components oscillating in opposite directions, there would be almost no compression or rarefaction of the liquid as such. Since a mechanical sound generator predominantly excites the compression and rarefaction waves of ordinary sound in helium II, the second sound turns out to be too weak to be detected.

However, these considerations suggested another way to detect the second sound waves. The opposed oscillations of the normal and superfluid components represent, in essence, oscillations of heat relative to the cold superfluid background. They should therefore cause oscillations of the liquid's temperature. It was natural to expect that such a "thermal wave" might also be radiated by a heater with oscillating temperature and that a heater of this kind might thus be made to serve as a second-sound generator. V. P. Peshkov of our institute performed the indicated experiments. They clearly confirmed the existence of second sound, in excellent agreement with the quantitative predictions of Landau's theory [*see chart at bottom of page 196*]. Second sound has since become one of the most important tools for investigation of helium II.

During the past 10 years Landau's theory has been developed further and has been supported by many experiments. Results obtained in a number of the world's low-temperature laboratories have generally been in splendid agreement with the theory. There remain, however, unanswered questions about superfluidity. Low-temperature physicists are particularly intrigued by the so-called "critical" phenomenon, which I have deliberately not mentioned hitherto. The point is that helium II does not actually exhibit superfluidity under any and all conditions. These properties are lost if the liquid is forced to move too rapidly (the critical speed depending strongly on the experimental conditions). Significant progress on this problem has been achieved during the past three years. The seed was planted by Lars Onsager of Yale University at a conference nearly 10 years ago in a remark on turbulence in the flow of fluids that was not particularly noticed at the time. The importance of this remark was later recognized by Richard P. Feynman of the California Institute of Technology, who then developed Onsager's idea. These authors presented convincing arguments for the notion that the critical

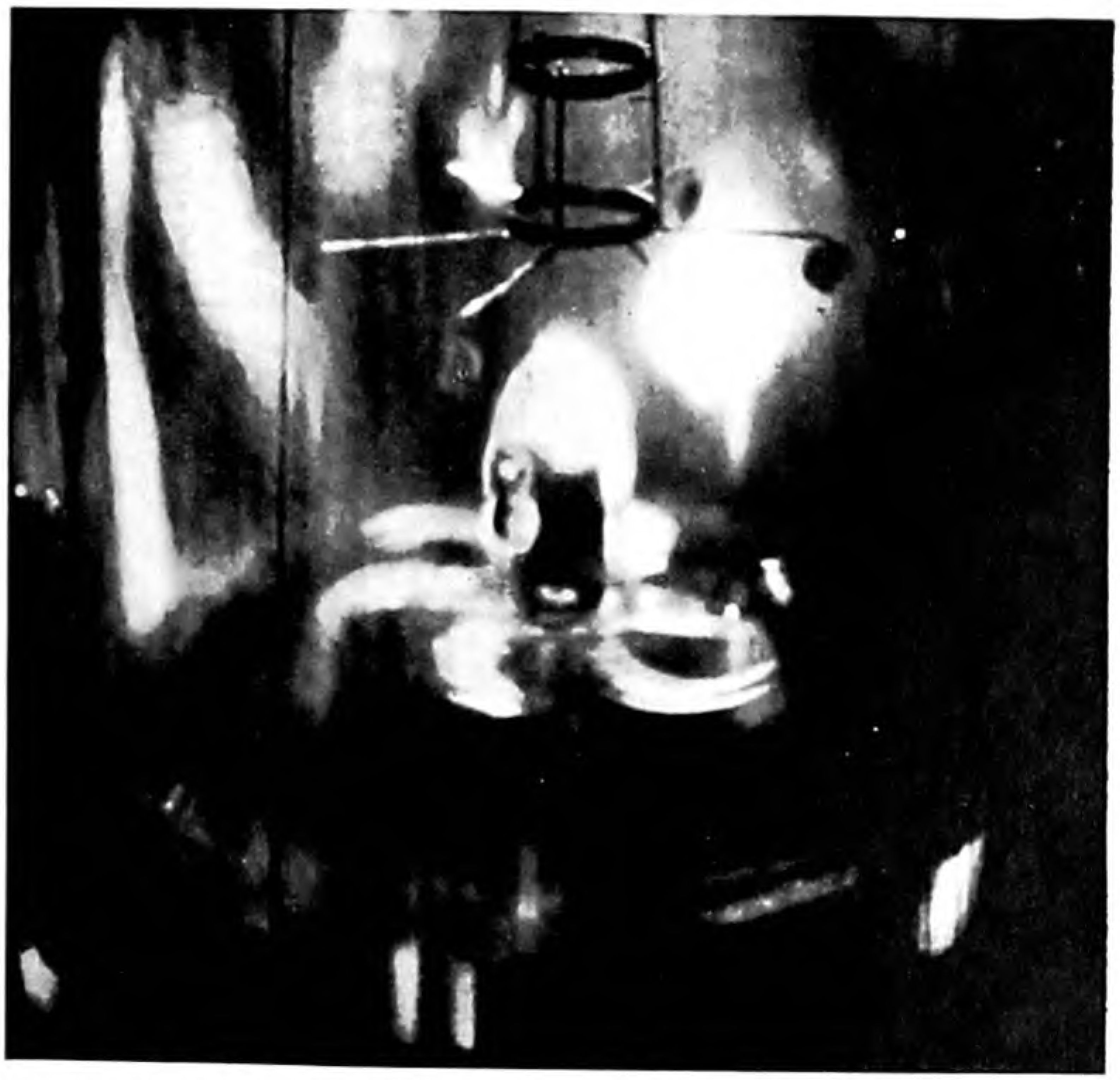
phenomenon is due to the formation in rapidly flowing helium II of a large number of microscopic vortices. In these vortices the superfluid component moves in a manner somewhat analogous to the way the air moves in a tornado. It would be difficult to detect such vortices directly, but H. E. Hall and W. F. Vinen at the Mond Laboratory in Cambridge have recently performed elegant experiments which give indirect evidence for the existence of the vortices, in particular the additional attenuation they induce in the second sound.

In concluding this short survey of the properties of helium II, I must return to the beginning of the article and correct a mistaken impression I may have created there. Helium is not a single substance; it exists in two stable isotopes, namely helium 4 and helium 3. As found in nature, helium is almost pure helium 4, and what has been said above refers to this isotope.

The enormous success of nuclear physics during the past decade has made it possible to obtain helium 3 in quantities sufficient for experimentation. In 1949 S. G. Sydoriak, E. R. Grilly and E. F. Hammel of the Los Alamos Scientific Laboratory showed that helium 3 liquefies at 3.2 degrees Kelvin, and a new "quantum liquid" was made available to physicists.

The isotopes of helium differ in an extremely important way which is related to the fact that the helium-4 nucleus contains an even number of particles (protons and neutrons), whereas that of helium 3 contains an odd number. This gives the two substances radically different quantum properties. (In physical terminology the helium-4 atom is said to obey Bose-Einstein statistics, whereas the helium-3 atoms obey Fermi-Dirac statistics.) There is every reason to believe that all liquids composed of "Bose-Einstein atoms" can become superfluid. As for liquids composed of "Fermi-Dirac atoms," there are in principle different possibilities. Only by experiment can we decide what we are dealing with in helium 3, the only Fermi-Dirac liquid in nature.

Experiments first performed by D. W. Osborne, B. Weinstock and B. M. Abraham at the Argonne National Laboratory have shown that liquid helium 3 does not become superfluid. The properties of this liquid thus lie outside the scope of an article on superfluidity. But I wish to note that in helium 3 physics has been provided with a "quantum liquid" of a new type. Though its properties are somewhat less spectacular than those of helium 4, they are no less interesting.



HELIUM TURBINE designed by Kapitza consists of small double-walled flask turned upside down and hung on the point of a needle in a helium bath. Helium flowing from inside the flask through the six bent capillary tubes causes the turbine to spin when light is shined upon it (*bottom*). This motion is stopped when the six vanes above the turbine are slipped down to oppose the reactive force of the outflowing helium. Inflowing helium II which replaces outflowing helium exerts no force, and the turbine remains motionless.

The Author

EUGENE M. LIFSHITZ is one of the leading theoretical physicists of the U.S.S.R. In 1958 he won the Lomonosov Prize of the Soviet Academy of Sciences for his theory of molecular attraction in solids. Now 43, Lifshitz graduated from the Kharkov Polytechnic Institute, then did research for several years at the Ukrainian Physical-Technical Institute before joining the staff of the Academy of Sciences in 1939. He is noted for his work in solid-state and low-temperature physics, and for his long collaboration with Lev D. Landau. One result of this collaboration is their seven-volume work *Quantum Mechanics: Course in Theoretical Physics,* which is now being published in English.

Bibliography

LIQUID HELIUM. Kurt Mendelssohn in *Handbuch der Physik,* Vol. 15, pages 370-461; 1957.

LOW TEMPERATURE PHYSICS: FOUR LECTURES. F. E. Simon, N. Kurti, J. F. Allen and K. Mendelssohn. Academic Press, Inc., 1952.

PROGRESS IN LOW TEMPERATURE PHYSICS, Vol. I. Edited by C. J. Gorter. Interscience Publishers, Inc., 1955.

SCIENTIFIC AMERICAN February 1957, Vol. 196, No. 2, pp. 71-82 OFFPRINT 225

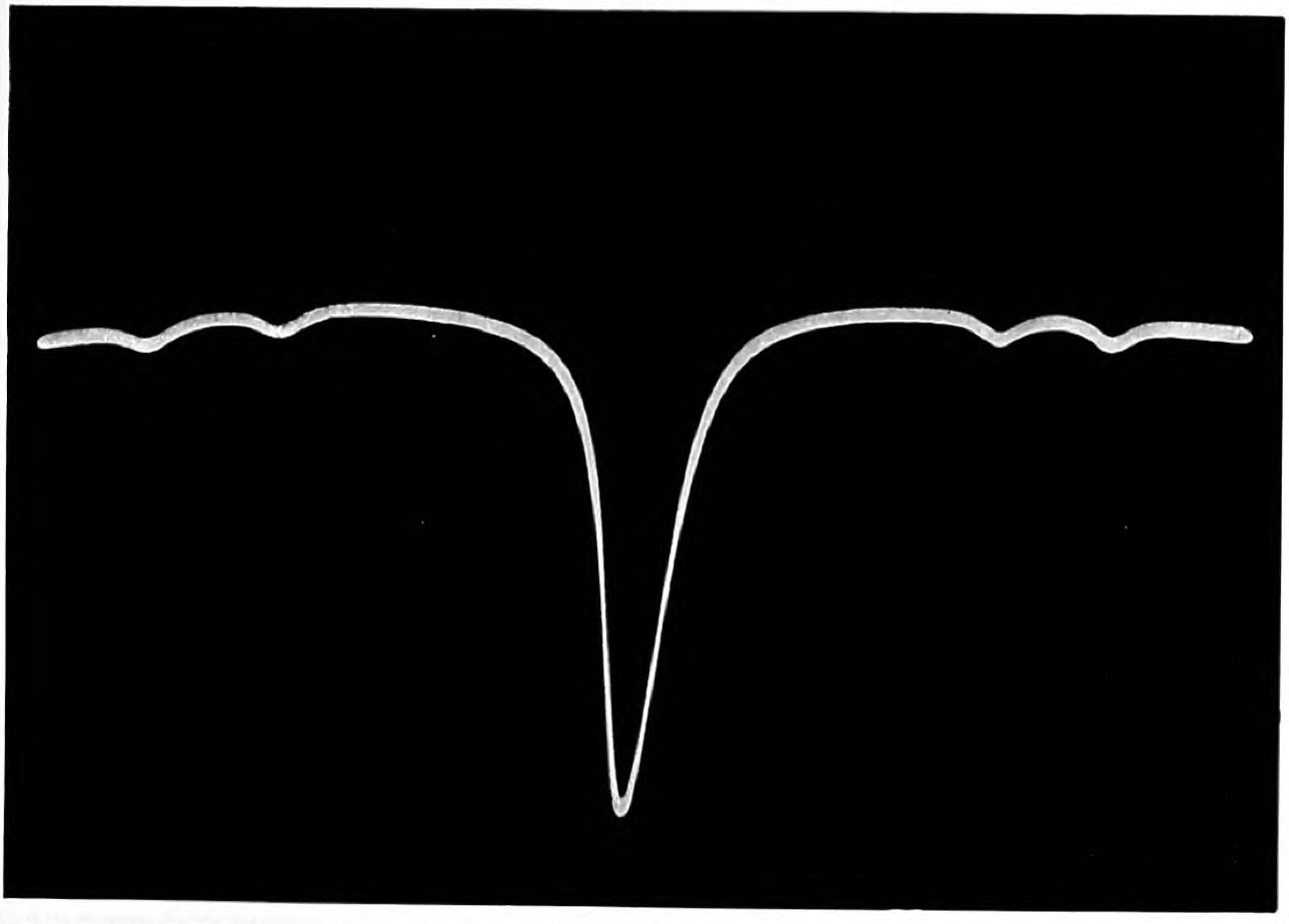
ATOMIC CLOCKS

by Harold Lyons

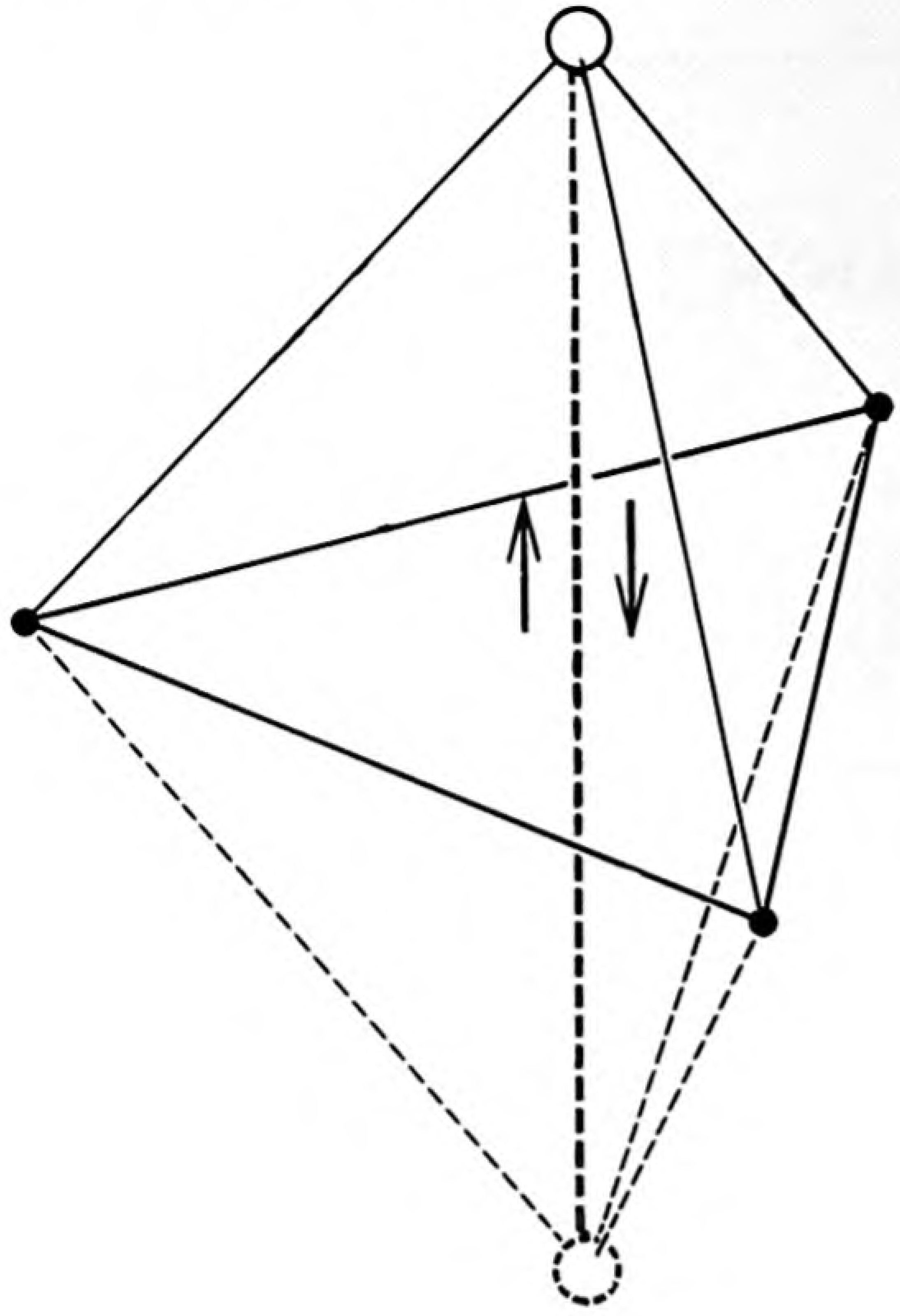
The "pendulums" which regulate them are the vibrating parts of atoms or molecules. So steady are these oscillations that atomic clocks keep better time than the spinning earth itself.

Philosophers and scientists in all ages have been fascinated by the mysteries of time—its relentless, arrow-like flight in one direction, its psychological vagaries, the difficulty of measuring it with absolute precision. In our atomic age the last of these aspects affords the most intriguing speculation and exploration. Because "the pendulum's swing is a variable thing," and the motions of the earth and stars are inconstant, today "the atom's vibrating has the highest rating" among chronologers. Let us then consider atomic clocks.

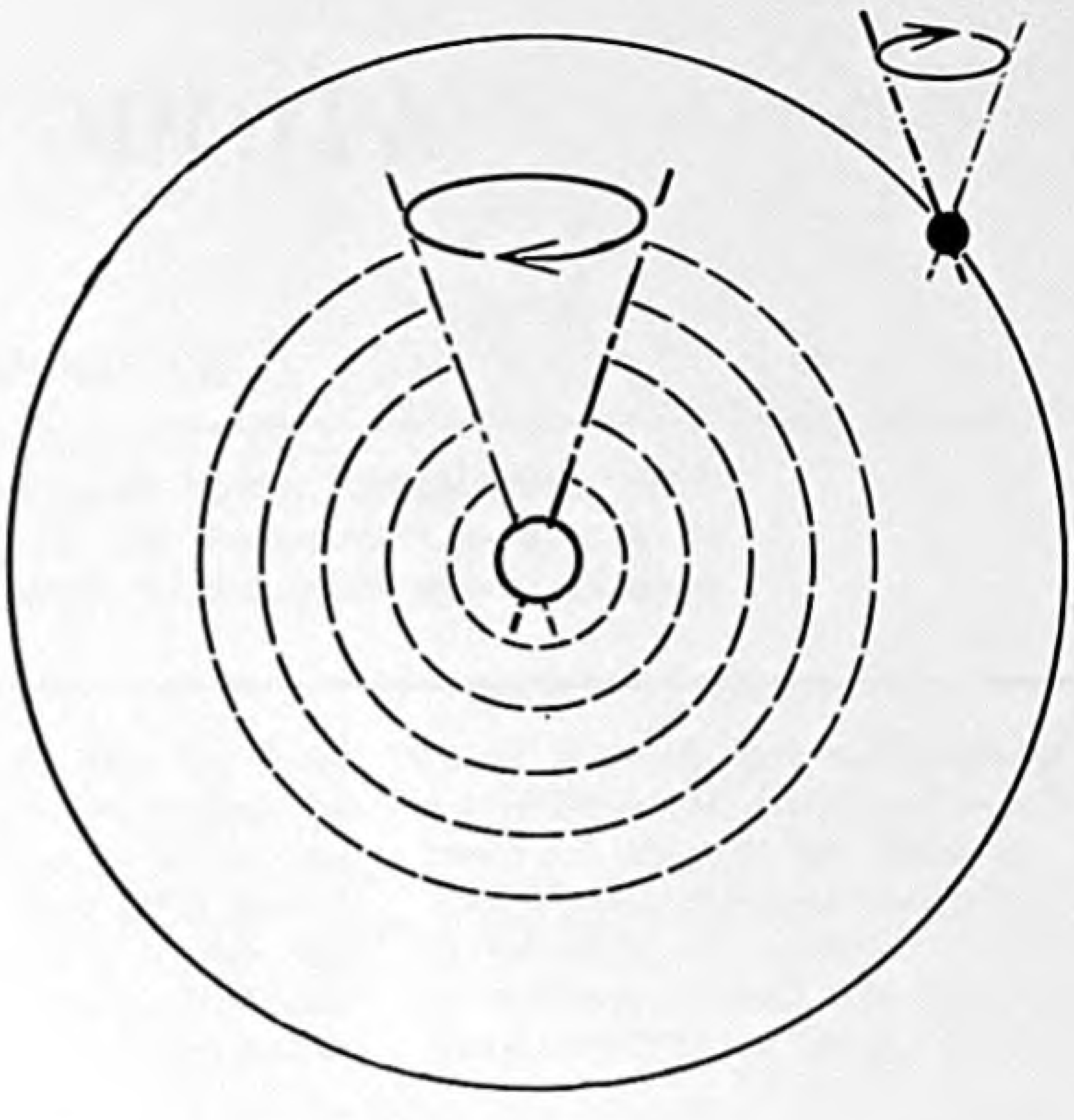
Most of us, when we ask for the right time, are satisfied with an answer accurate to a few seconds or so. For the "split-second" timing of a race, tenths of seconds will do. But in many areas of modern science and technology the question of the right time enters a different realm.



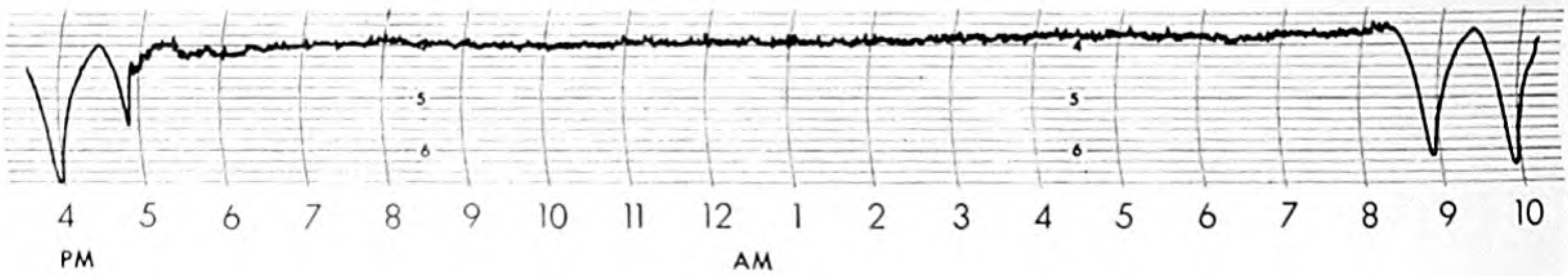
ABSORPTION CURVE of ammonia is recorded on an oscilloscope. The trace shows power received from a beam of radio waves transmitted through ammonia gas. Frequency varies along the horizontal axis. At the resonant frequency most of the wave energy is absorbed, as is shown by the dip in the curve. The range of frequencies indicated by the dip limits the accuracy of ammonia clocks.



AMMONIA MOLECULE has the shape of a pyramid. Hydrogen atoms (*black dots*) form a triangular base. Nitrogen atom (*open circle*) is at apex. It can oscillate between positions above and below the base, traveling along the path marked by colored line.



CESIUM ATOM has a single electron (*black dot*) outside of a number of filled electron shells (*broken circles*). The electron and nucleus are spinning magnets; each wobbles on its axis, as is indicated by colored arrow. Wobble is the ticking of a cesium clock.



STABLE PERFORMANCE of an ammonia clock over a 15-hour period is demonstrated by a record of its frequency changes. The frequency is measured on the vertical scale of the chart, each small division corresponding to a change of less than one part in 10 million. During the period from 5 p.m. to 8 a.m. the quartz crystal oscillator was locked to the frequency of the ammonia molecules.

In the laboratory we must deal with thousandths, millionths, even billionths of a second.

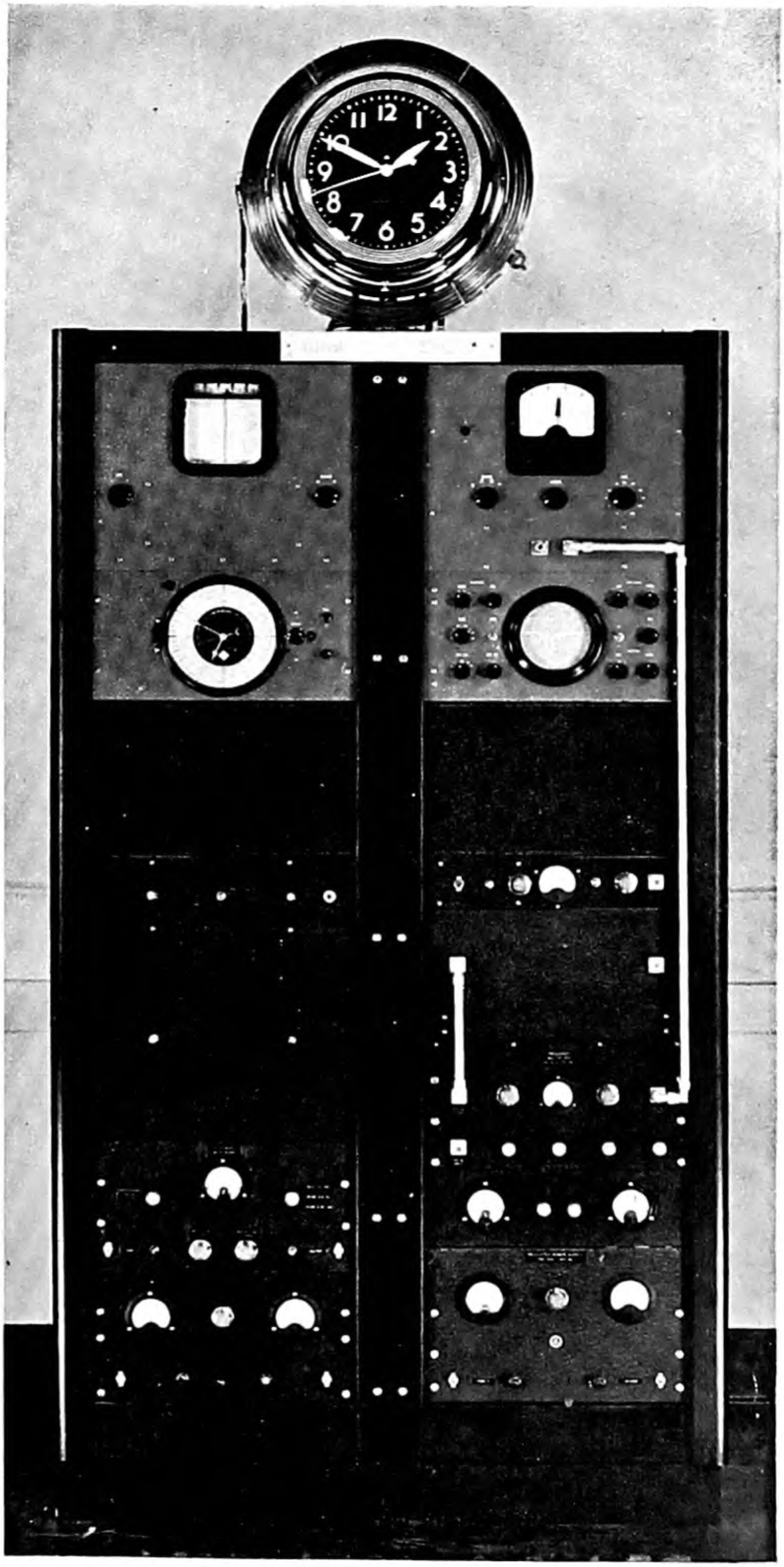
The measurement of any physical quantity reduces in the last analysis to a matter of counting units. To find the distance between two points, for example, we choose some convenient yardstick and count the number of times it can be laid end to end from one point to the other. To find the elapsed time between two instants we choose a convenient unit, such as the time required by a certain pendulum to complete one swing, and count the number of swings in the interval. However, the swings of a pendulum are not precisely the same from one to the next. The central problem of exact time measurement is to find some periodic cycle that never changes, or changes so little that the variation can be disregarded. For ages immemorial we have reckoned time by the rotation of the earth relative to the stars. Now we have begun to seek more precise standards in the tiny world of molecules and atoms. There we find processes whose regularity makes it possible to measure time with undreamed-of accuracy.

The Clock on the Wall

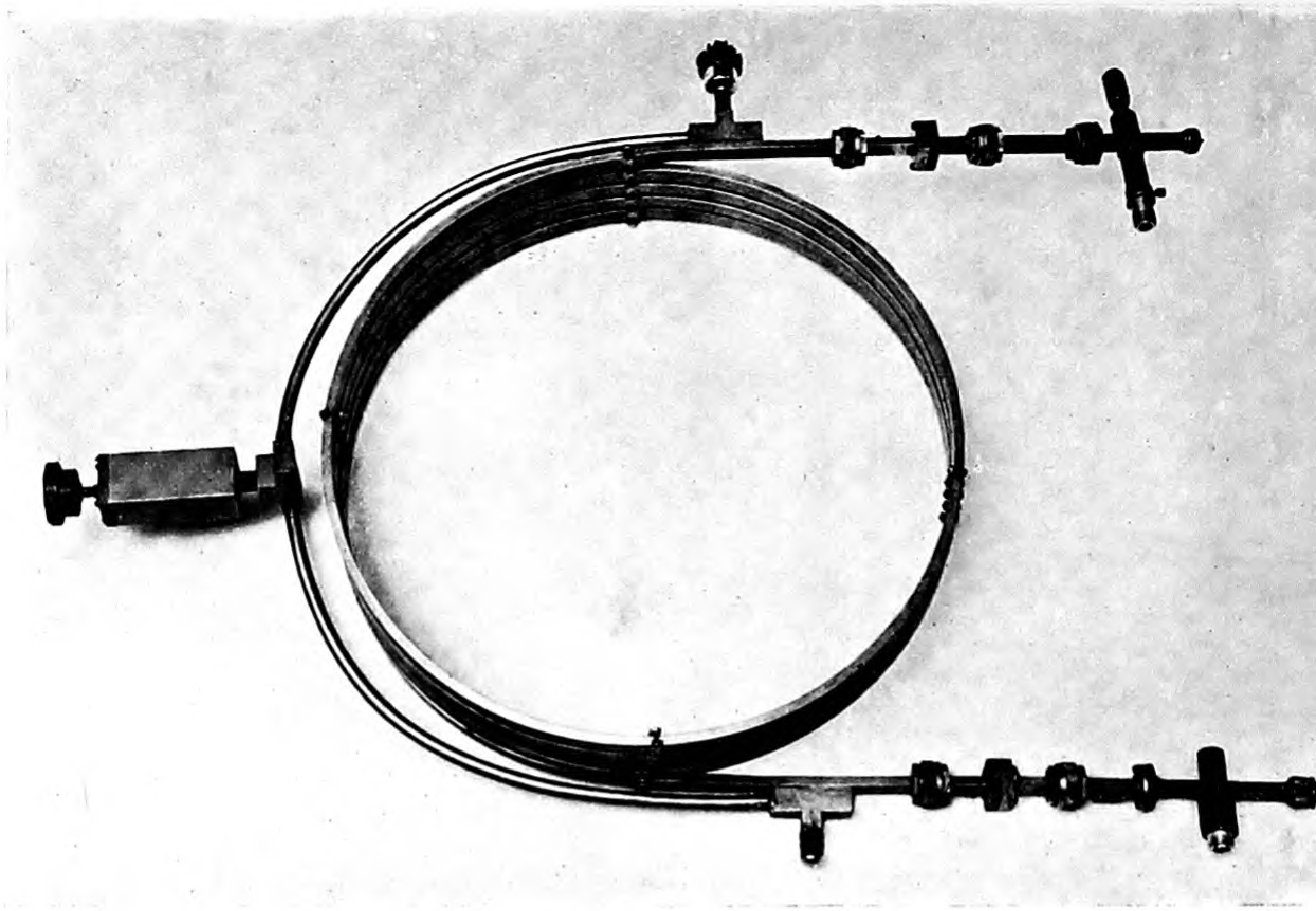
Before looking into these cosmic clocks in more detail, let us consider briefly how ordinary clocks operate. In the household electric clock the "pendulum" is the cycle of the alternating current. Hence the accuracy of the clock depends on the steadiness of the rate of alternation of the current. For household purposes the 60-cycle rate maintained by the power-generating station is steady enough.

For higher precision, laboratories and observatories use quartz crystal clocks. Here a quartz crystal controls the frequency of an electronic oscillator, such as is used in radio broadcasting. A crystal of quartz, when subjected to an alternating electric field, tends to vibrate at its own specific, sharply defined rate. Placed in an oscillator circuit, the crystal imposes its steady natural frequency on the circuit. The resulting current can run a synchronous clock motor with an error of no more than one part in a billion or so, depending on the length of the interval involved. However, changes in temperature and other conditions produce tiny shifts in the crystal frequency, and as a crystal ages its frequency drifts.

Fundamentally all man-made clocks are set by some master clock in nature, which at present is the 24-hour rotation of the earth. The complete rotation is



FIRST AMMONIA CLOCK was completed at the National Bureau of Standards in 1949. The wave guide that contains the ammonia gas can be seen wound around the face of the electric clock. The cabinet below houses crystal oscillator and other electronic circuits.



ABSORPTION CHAMBER for the ammonia clock is a hollow, rectangular, spiral wave guide which contains ammonia gas at low pressure. A radio signal is fed in at one end and detected at the other. When the frequency of the wave falls within the absorption

timed precisely by recording the instant a point on the earth passes under a chosen star in the sky on successive nights. The interval is divided into 86,400 parts, which gives the length of a second.

But in the computation of the length of the day, corrections have to be made for a number of irregularities, including wobbles in the earth's rotation on its axis. When all the corrections have been made, an insurmountable uncertainty still remains: the rate of rotation of the earth itself fluctuates unpredictably. So in the end there is an irreducible variation which can be as large as one part in 20 million.

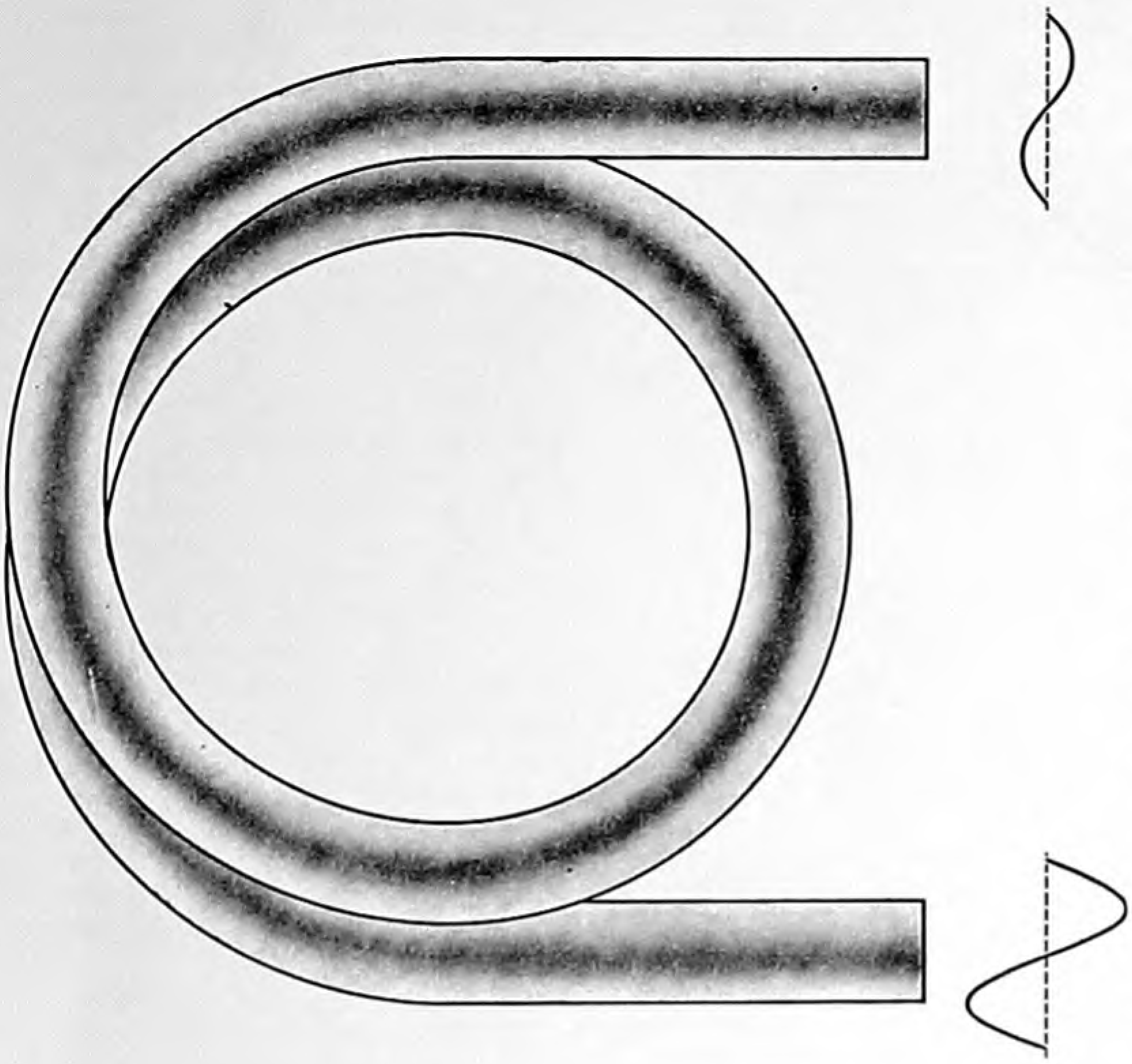
All this explains why so much effort is being devoted to finding clocks which will keep better time than the earth and stars. The atomic clocks offer great advantages. The motions of atoms and molecules, which can serve as "pendulums," are absolutely pure and regular. Their rates are inexorably fixed by the laws of the atomic world.

Some of the motions in the atomic world—*e.g.*, the vibrations of electrons that radiate visible light—are much too rapid to be counted. But there are atomic oscillations in the radio microwave region, with frequencies in the range of a few billion cycles per second, which can be counted accurately by present-day techniques and equipment.

The Ammonia Clock

The first atomic clock devised is the one based on vibrations of the ammonia molecule [see "Radio Waves and Matter," by Harry M. Davis; SCIENTIFIC AMERICAN, September, 1948]. This molecule, made up of three hydrogen atoms and one nitrogen atom, has the shape of a pyramid. The hydrogens are at the corners of the triangular base, and the nitrogen is at the apex [*see diagram at top of page 202*]. According to the rules of classical physics, the forces between the atoms should hold the nitrogen in place at the top of the pyramid. But experiments have shown that the nitrogen can actually plunge down through the triangular base and come out to an apex position on the other side—a phenomenon which can be explained by quantum mechanics. The motion is merely hindered, not prevented, by the interatomic forces. And of course if the nitrogen can pass through in one direction it can also reverse its path; in other words, it can vibrate up and down through the base. As we should expect from quantum theory, the vibration can take place only at a sharply defined frequency, which happens to be 23,870 megacycles.

Whenever it is excited by a sufficient amount of energy, the ammonia molecule starts to vibrate with its characteristic frequency. It is like a pendulum which is set swinging by a push. If the push is supplied rhythmically, and in time with the natural frequency of the pendulum, the resulting swing is much more vigorous. That is, the molecule absorbs more energy from the source of supply and converts it into the energy of its own oscillator. A radio wave at a frequency of 23,870 megacycles makes the



band of the ammonia molecule the output signal is sharply reduced. At right is a schematic diagram of the unit showing the input signal at the bottom end and the output at the top.

nitrogen atom absorb large quantities of energy and vibrate strongly.

In 1948 a group of workers at the National Bureau of Standards built an ammonia clock. Their design contains essentially two pendulums: a quartz crystal and a collection of ammonia molecules. The ammonia serves to correct small errors or irregularities in the crystal-controlled oscillator. The oscillator in turn runs an ordinary synchronous electric motor like the one in a kitchen clock. For this purpose its frequency has to be reduced to that of an alternating electric current suitable for running the motor–*i.e.*, to the neighborhood of 60 cycles per second. The crystal frequency is cut down to a precise fraction of the original by means of electronic circuits analogous to a train of gears which converts the rapid rotation of a small gear to the slower rotation of a large one.

The ammonia clock works as follows. First the quartz crystal is set vibrating at a frequency which, when multiplied electrically, yields a frequency close to that of the ammonia molecule. These rapid oscillations are then converted into radio waves by means of a small antenna and are fed into a long chamber, or wave guide, containing ammonia gas. If the oscillations happen to be at the same frequency as that of the ammonia molecule, most of the radio energy will be absorbed by the ammonia, and little will get through the chamber to the other end. But if the two frequencies do not quite agree, most of the radio energy will pass through the chamber to a receiver at the far end. The receiver acts as a feedback mechanism, feeding a servomotor which adjusts the frequency of the oscillator circuit to agree with the ammonia vibration rate.

The feedback circuit has a time lag, and its corrections are not absolutely exact. The vibration of the molecules themselves exhibits an inherent fuzziness. In an assembly of ammonia molecules there is some spread of the rates of vibration. There are two chief reasons for this. First, the moving molecules of ammonia gas constantly collide with one another and with the walls of the chamber. At every collision the atoms of the ammonia molecule are subjected to outside forces which slightly spread the molecule's frequency from its normal value. The second factor in shifting the frequency is a Doppler effect. To radio waves passing through the chamber, the vibration frequency of ammonia molecules moving away from the waves appears to be slightly lower than it actually is, and the frequency of molecules moving toward the waves appears to be higher. Both the collision and Doppler shifts are sufficient to give a measurable spread around the central frequency and thus to limit the possible accuracy of the ammonia clock.

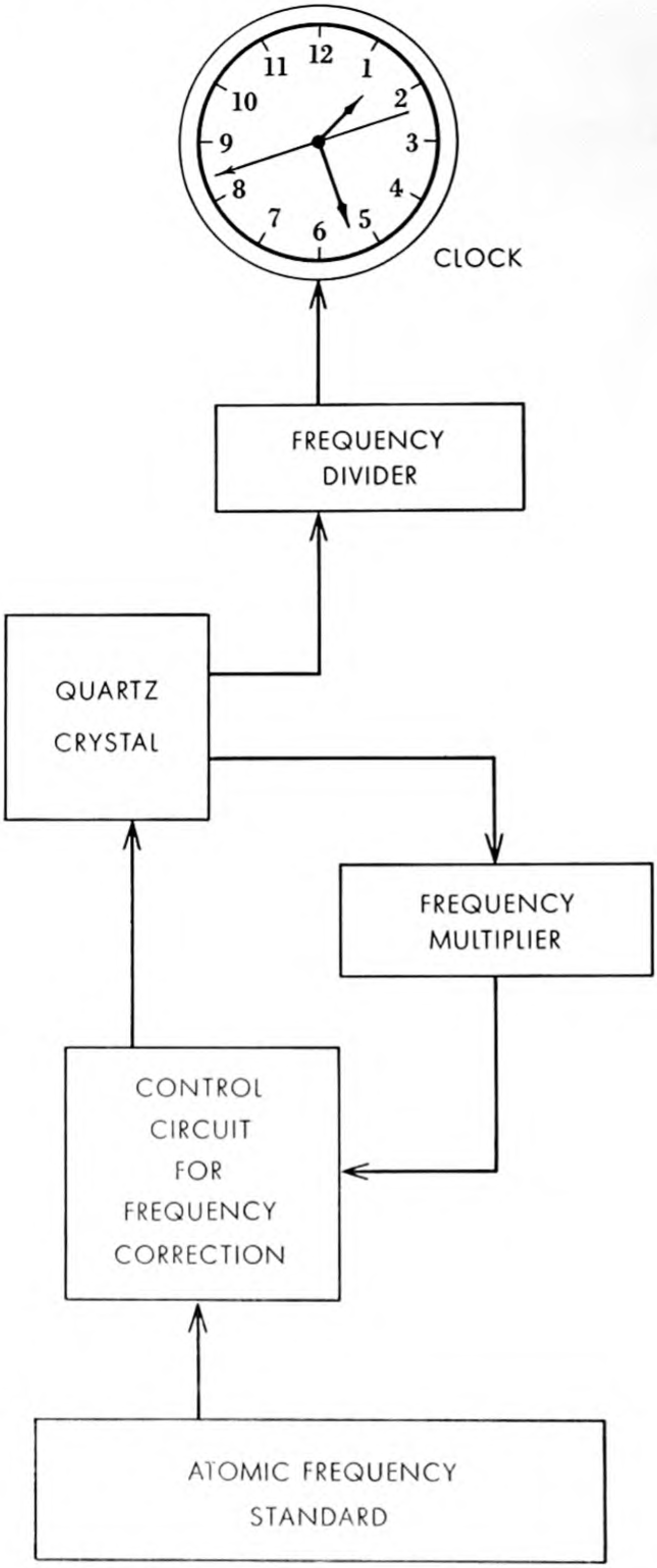
The accuracies that have been obtained are quite impressive, however. An improved version of the original Bureau of Standards ammonia clock is stable to within one part in 100 million. J. Rossel of the Swiss Laboratory for Time-Keeping Research has reported that a newer ammonia clock built there can be held steady up to two parts in a billion. K. Shimoda of Japan also has designed a new form of ammonia clock which can control frequency to two or three parts in a billion.

The Cesium Clock

Now an atomic clock of considerably greater precision–the most accurate yet built–has been made with cesium, a silvery metal which is liquid at room temperature. It was designed by the Bureau of Standards group who made the first ammonia clock.

The cesium atom, like the ammonia molecule, has a natural vibration whose frequency is in the microwave region. Its frequency is 9,192 megacycles. This puts it, very conveniently, in the range of three-centimeter microwaves, a region which has been intensively exploited for radar work. Thus the necessary equipment and techniques are ready to hand.

What can go on in the cesium atom at this comparatively leisurely pace? It turns out to be a magnetic process. Cesium is an alkali metal, which means that outside its filled electron shells it has a single outermost electron whose spin makes it a magnet. (The magnetisms of the electrons in the closed shells do not count, because they cancel one another.) The spinning nucleus of the cesium atom also is a magnet. Thus the atom contains two small, spinning magnets, each in the force field of the other. Neither magnet maintains a rigidly fixed direction. They are both like tops spinning in a gravita-



CONTROL CIRCUITS for an atomic clock are shown in this schematic diagram. Part of the output of a quartz crystal oscillator is reduced by a frequency dividing circuit to about 60 cycles per second and fed into an ordinary electric clock. Another part of the output is multiplied to the atomic vibration frequency and fed to a circuit which compares it with the atomic frequency itself. Any difference is translated into an electric signal which feeds back to the oscillator and brings its frequency into agreement with the atomic standard.

tional field. That is, they wobble or precess around a fixed line [*see diagram on page 202*]. The rate of precession is 9, 192 megacycles per second. This represents, in effect, the ticking of the cesium clock.

If cesium atoms are placed in an electromagnetic field which oscillates at 9,-192 megacycles, the electrons can absorb or emit energy and flip over to a different orientation. This change of energy state is the mechanism of the clock.

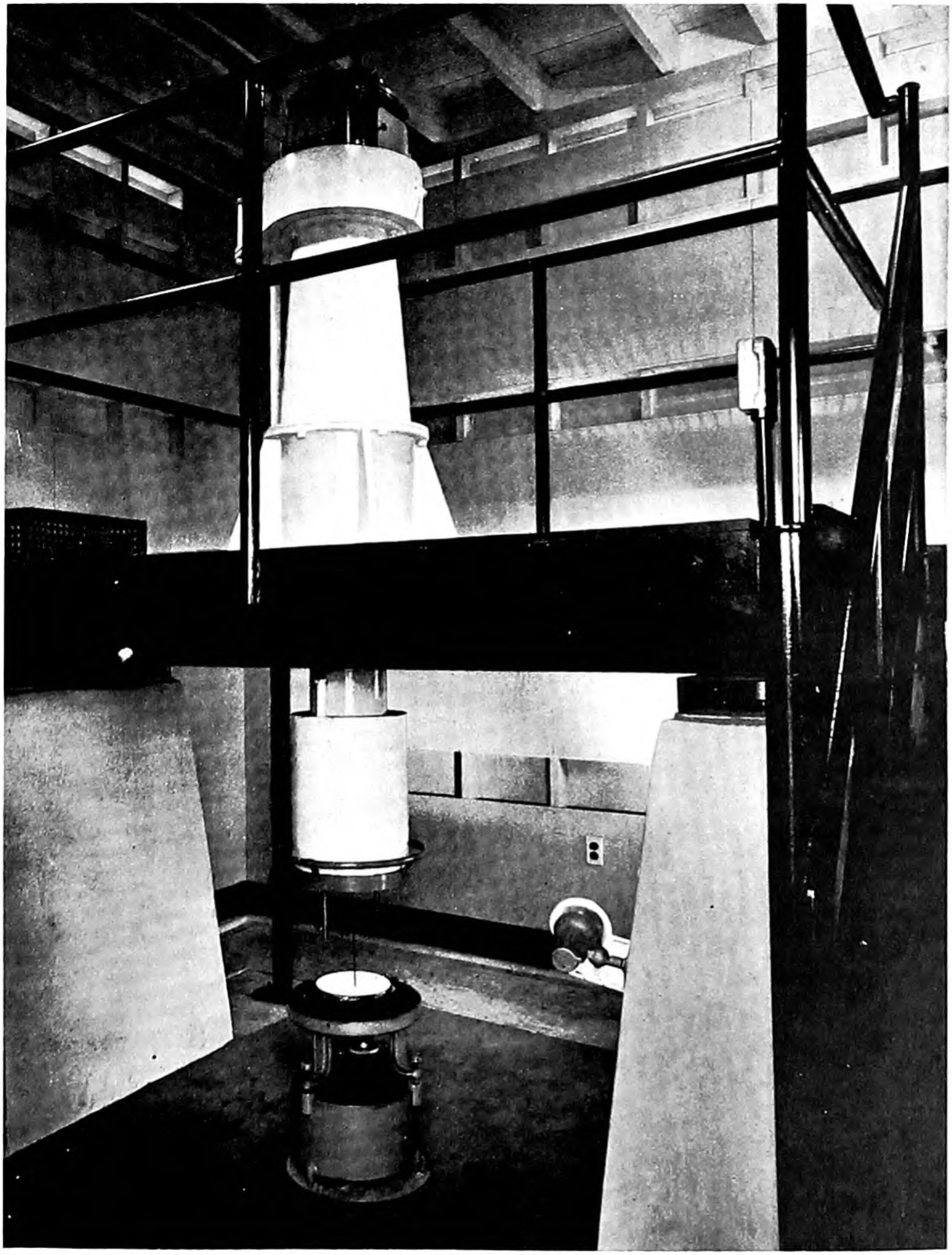
For the cesium clock the element is heated to the gaseous state in an electric furnace, and the cesium atoms are discharged through a small opening into a long, evacuated tube. They travel down the tube in a beam like a file of marching soldiers, thus avoiding collisions with one another. The beam is sent through a magnetic field which acts to select only atoms in certain energy states. Next the beam passes through a section where it is exposed to radio waves at the frequency of 9,192 megacycles. Finally it passes through a second magnet just like the first and then approaches a detector wire where the cesium atoms will produce a current if they hit it [*see diagram at top of page 210*].

If the radio field is on the correct frequency, large numbers of atoms in the beam change their energy. The second magnet now deflects these atoms so that they reach the detector. On the other hand, if the radio frequency does not agree with the natural frequency of the cesium atoms, they pass through the second magnet and are deflected so that few reach the receiver. As in the ammonia clock, the receiver actuates a mechanism which adjusts the oscillator frequency so as to keep the received current at a maximum.

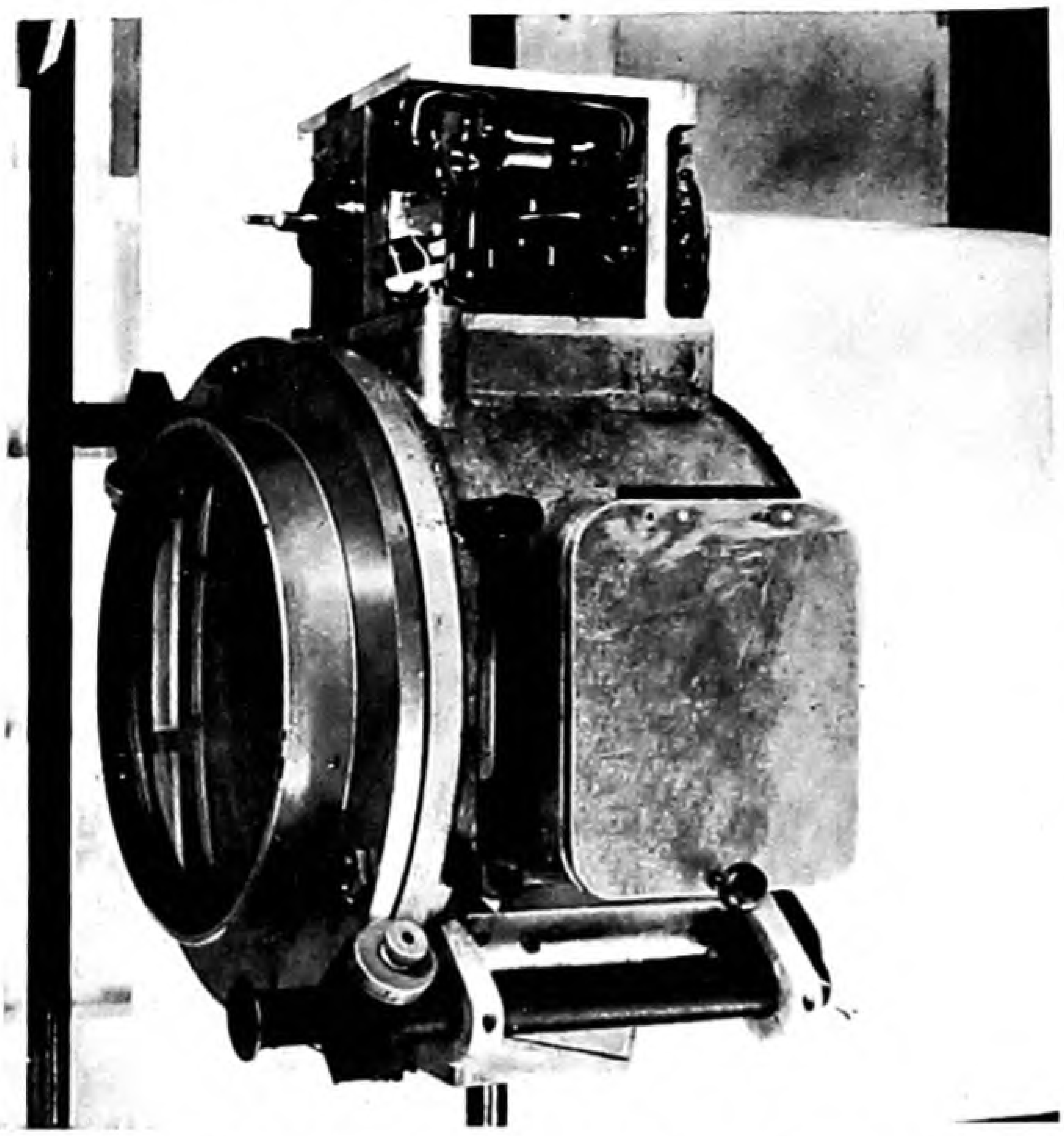
The cesium clock is extremely accurate because the spectrum line is very sharp. The device eliminates collisions between atoms and the Doppler effect, which broaden the absorption band in the ammonia clock. There is no Doppler effect because the radio waves attack the passing beam at right angles instead of moving along the same line of travel.

With precise control of the radio frequency it is possible to achieve accuracy to at least one part in 10 billion in the cesium clock. This would correspond to an error in timekeeping of one second in 300 years!

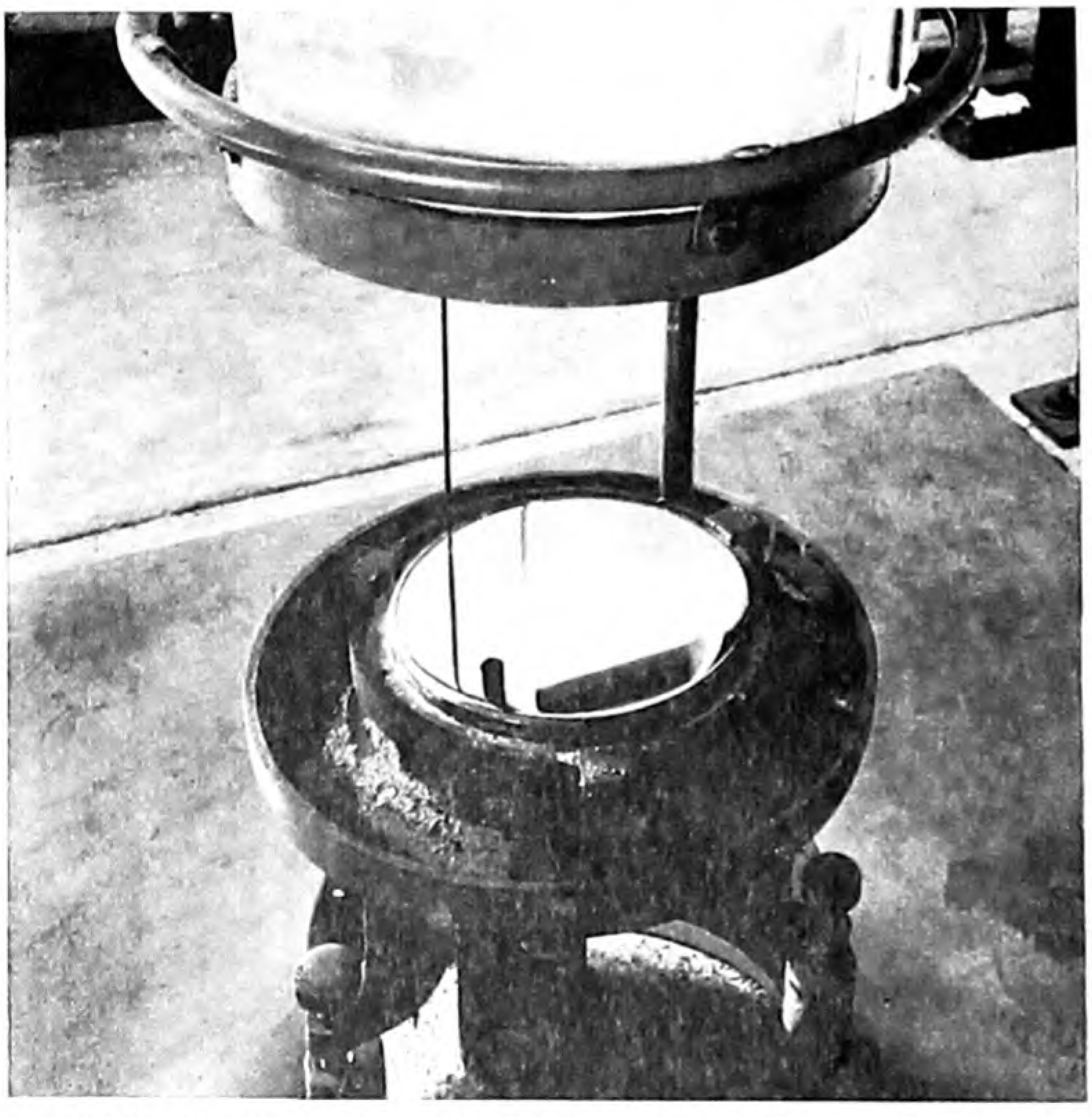
Already a cesium clock at the National Physical Laboratory in England has operated to an accuracy of one part in one billion. L. Essen and J. V. L. Parry have determined the center frequency of the cesium spectrum line in terms of a pro-



PHOTOGRAPHIC ZENITH TUBE makes pictures of stars as they pass directly overhead. The time between transits on successive nights determines the length of the solar day. The tube is fixed; below it is a basin of mercury which acts as a mirror. Because the mercury is liquid, it is absolutely level. The mirror reflects the light of stars to a photographic plate within the tube (*see picture at top of next page*). This instrument is at the U. S. Naval Observatory in Washington, D.C. A similar instrument is at Richmond, Fla.



LENS of photographic zenith tube has an aperture of eight inches. It is shown mounted on a housing which contains the photographic plate. The motor at top moves the plate to keep star image tracked for 20-second exposure. About 15 stars are photographed a night.



MERCURY MIRROR provides a flat and truly level surface. It is raised or lowered to focus the star image on the photographic plate under the lens. Above the surface of the mercury is a rod; its reflection can also be seen in the mercury. When the tip of this rod just touches the surface of the mercury, the image is in exact focus at the photographic plate.

visional, uniform time scale estimated by the Greenwich Observatory. Their answer is 9,192,631,830 cycles per second, with a possible error of 10 cycles. A cesium clock of at least this accuracy has also been built by the National Company in the U. S.

An ingenious scheme to improve the cesium clock's potential accuracy is being explored by Jerrold R. Zacharias and his group at the Massachusetts Institute of Technology. Zacharias plans to increase the length of exposure of the cesium beam to the radio waves, which will sharpen the absorption line. He proposes to do this by shooting the cesium atoms upward and exciting them with the radio field near the top of their trajectory, when they are about to fall and are moving slowly. His plan may give the cesium clock an accuracy to one part in 1,000 billion or better.

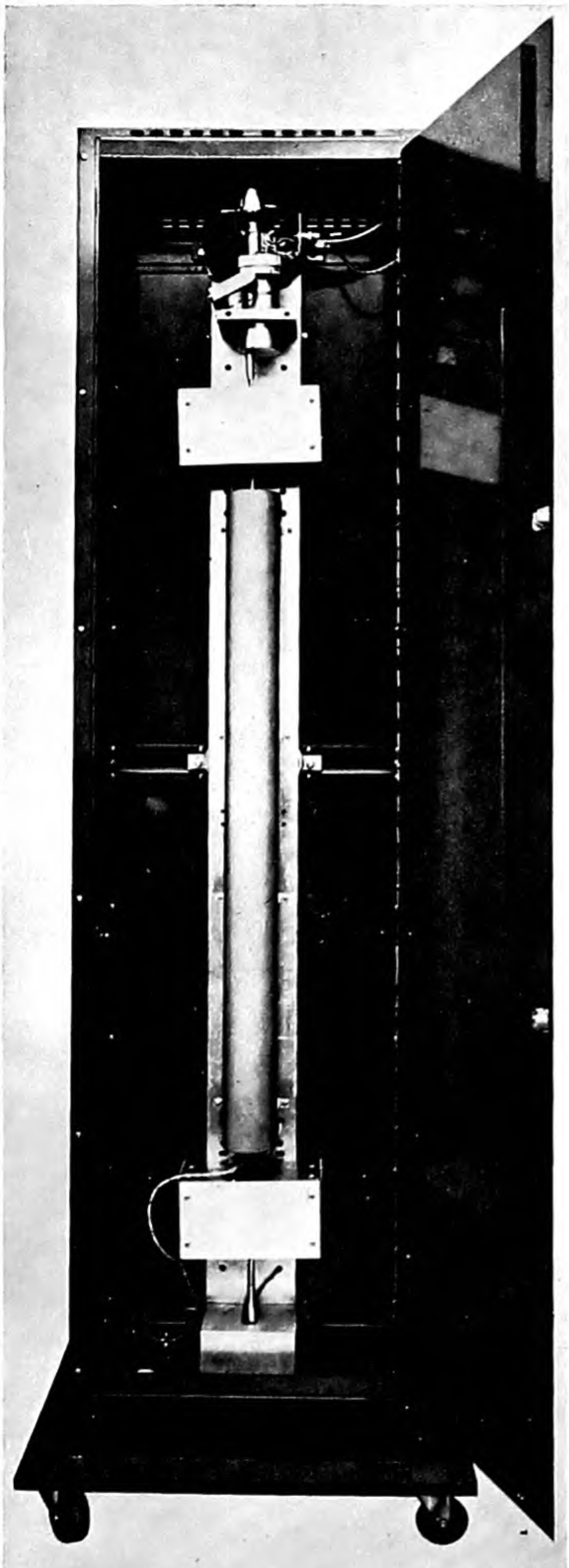
The Maser Clock

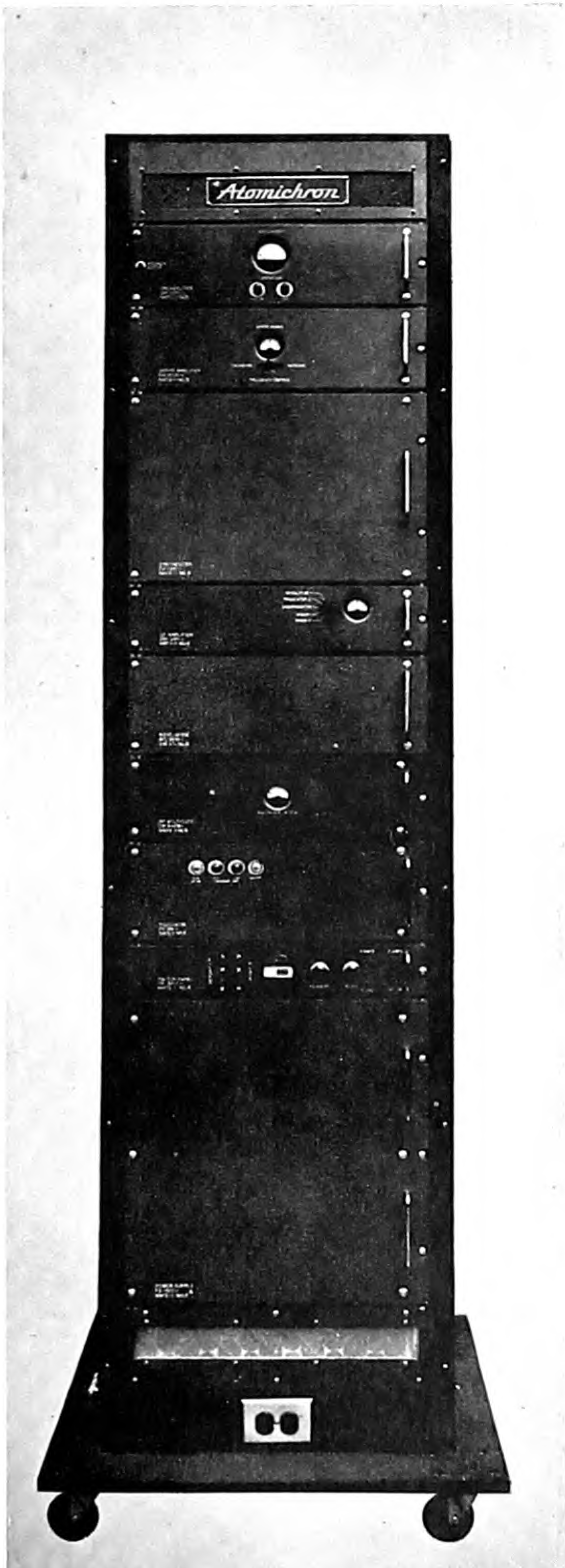
Why not tell time directly from an atom's own vibrations, instead of by the roundabout method of seeking its absorption frequency? The idea is indeed feasible, and a new atomic clock based upon it has been developed by C. H. Townes, J. P. Gordon and H. J. Zeiger at Columbia University. They call it the "maser" (for "microwave amplification by stimulated emission of radiation"). Their timekeeper is the ammonia molecule in the excited state, in which it emits rather than absorbs energy.

A beam of ammonia gas molecules coming from a high-pressure bottle enters a tube where it is first subjected to an electric field. This field acts as a focuser that disperses molecules in the low-energy, absorbing state and concentrates the emitters. The beam of emitters then flows into a "cavity resonator," where the molecules radiate their microwave energy. The size of this cavity is adjusted so that it resonates at precisely the frequency of the ammonia's radiation. Thus the energy emitted by the molecule is reinforced, and a strong oscillation is set up. The oscillation can be used to control the synchronous motor of an electric clock by means of a servomechanism.

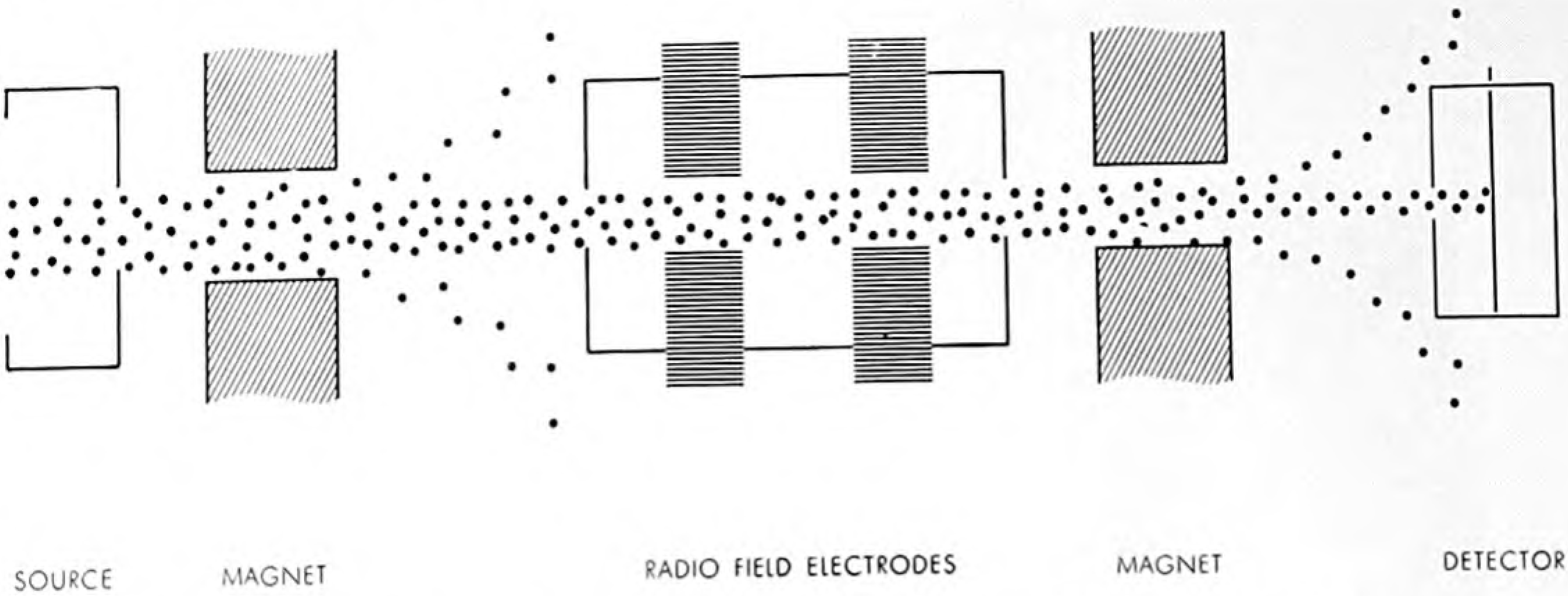
The maser has produced the purest oscillations ever generated—a signal very close to a single frequency. The frequency was stable to one part in 10 billion or better for more than an hour. This performance can probably be improved.

If the number of molecules fed into the maser's resonator is reduced below the level needed to sustain oscillations, and a small amount of energy in the





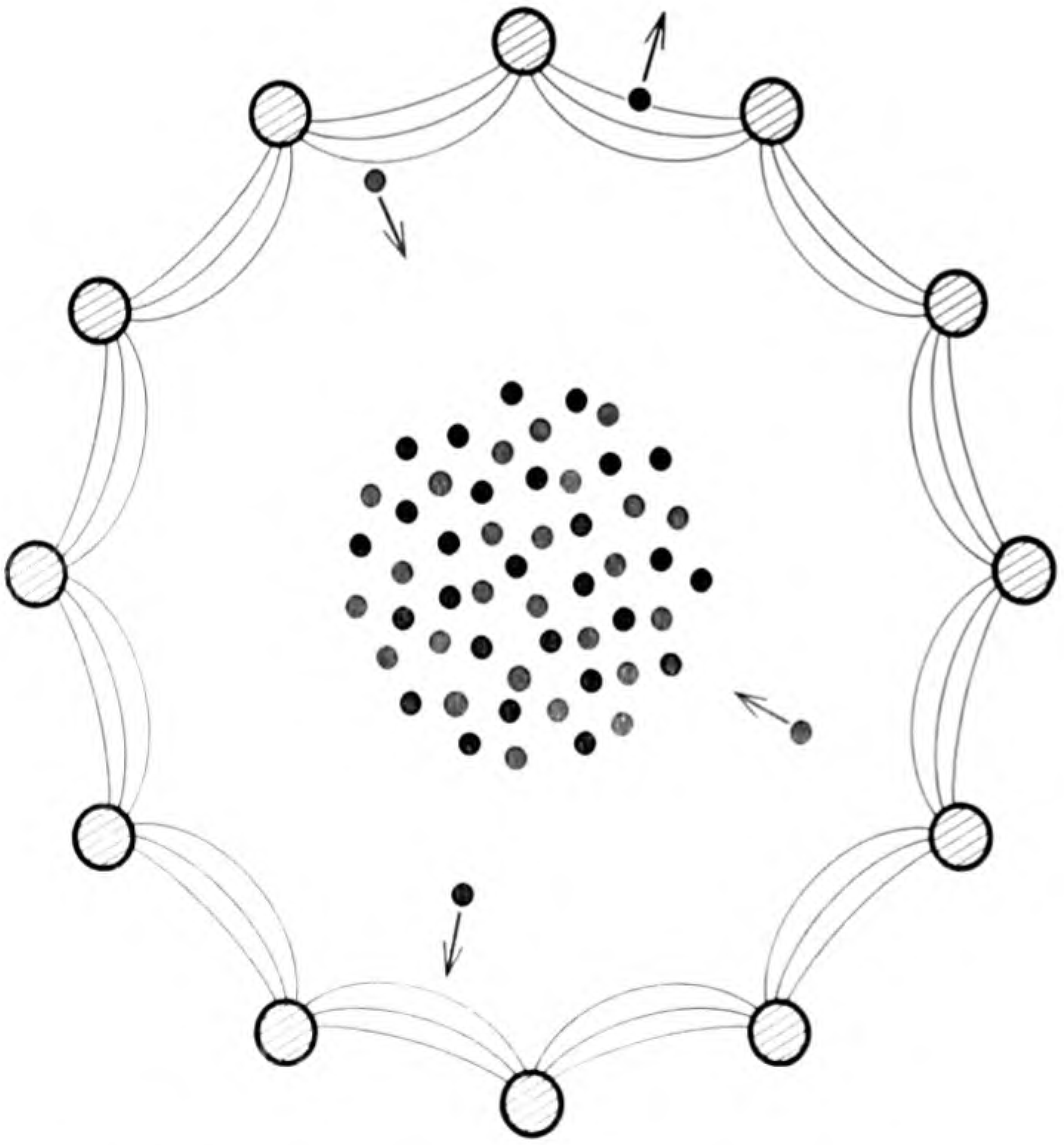
CESIUM BEAM EQUIPMENT at the Naval Research Laboratory in Washington is seen from the back at left and front at right. Cesium atoms are injected at the bottom and travel up the vertical tube to target chamber at top. Deflecting magnets are behind the two rectangular plates. This device is made by the National Company. It is the first commercial atomically controlled oscillator.



CESIUM BEAM PRINCIPLE is illustrated above. The beam emerging from the source (an electric furnace) contains atoms in two energy states. These are deflected by the first and second magnets in such a way that they miss the detector wire at right. When the atoms are excited by the radio field during their passage between the two magnets, they make transitions between the two energy states. The second magnet now deflects them in opposite direction so that they either land on detector wire or are refocused.

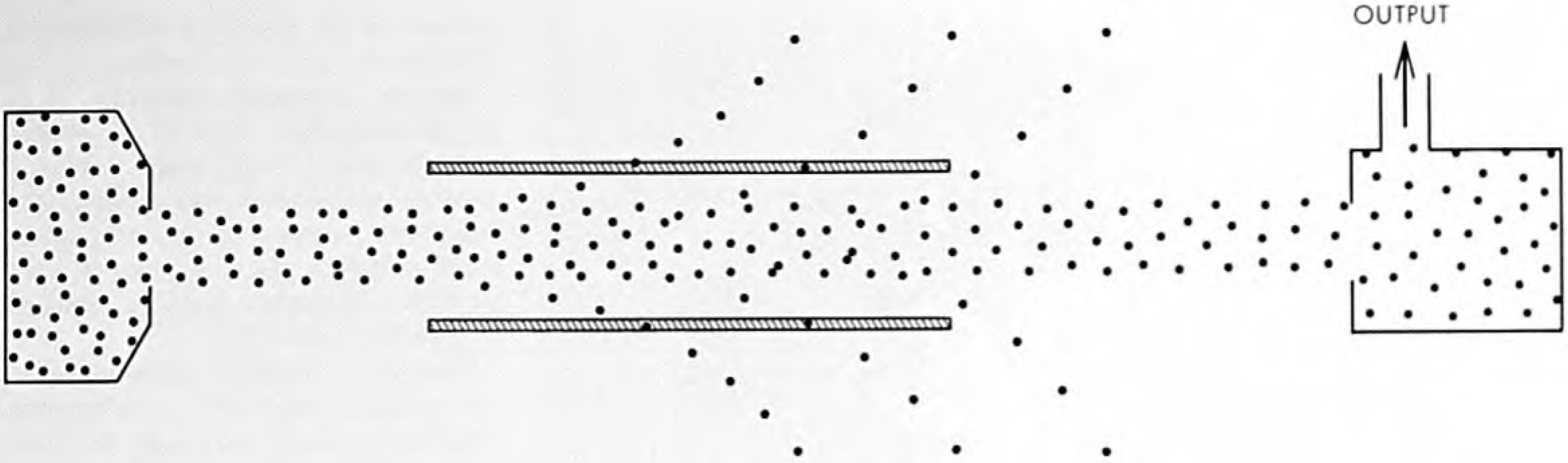
form of a radio wave of the right frequency is then added, the vibrations of the ammonia molecules will amplify the input signal. In this form the maser is an exquisitely selective and noiseless amplifier. It produces strong amplification when a weak signal at the proper frequency is fed into the cavity. Even if the input signal is contaminated with other frequencies, the ammonia responds only to its own vibration rate, so that its tuning is very selective.

A number of laboratories are now building masers and applying them in many areas of research. Some experimenters are testing new designs which may give even higher degrees of accuracy. R. H. Dicke of Princeton University proposes to use rubidium atoms instead of ammonia molecules and to include argon gas as a buffer to insulate the rubidium atoms against collisions with one another. D. D. Babb, formerly of the Signal Corps Engineering Labora-



MASER FOCUSER seen end-on shows rods as hatched circles and the electric field as black curved lines. The colored dots are emitting molecules; black dots, absorbers.

MASER at Columbia University is seen with the front plate of the vacuum chamber removed. The



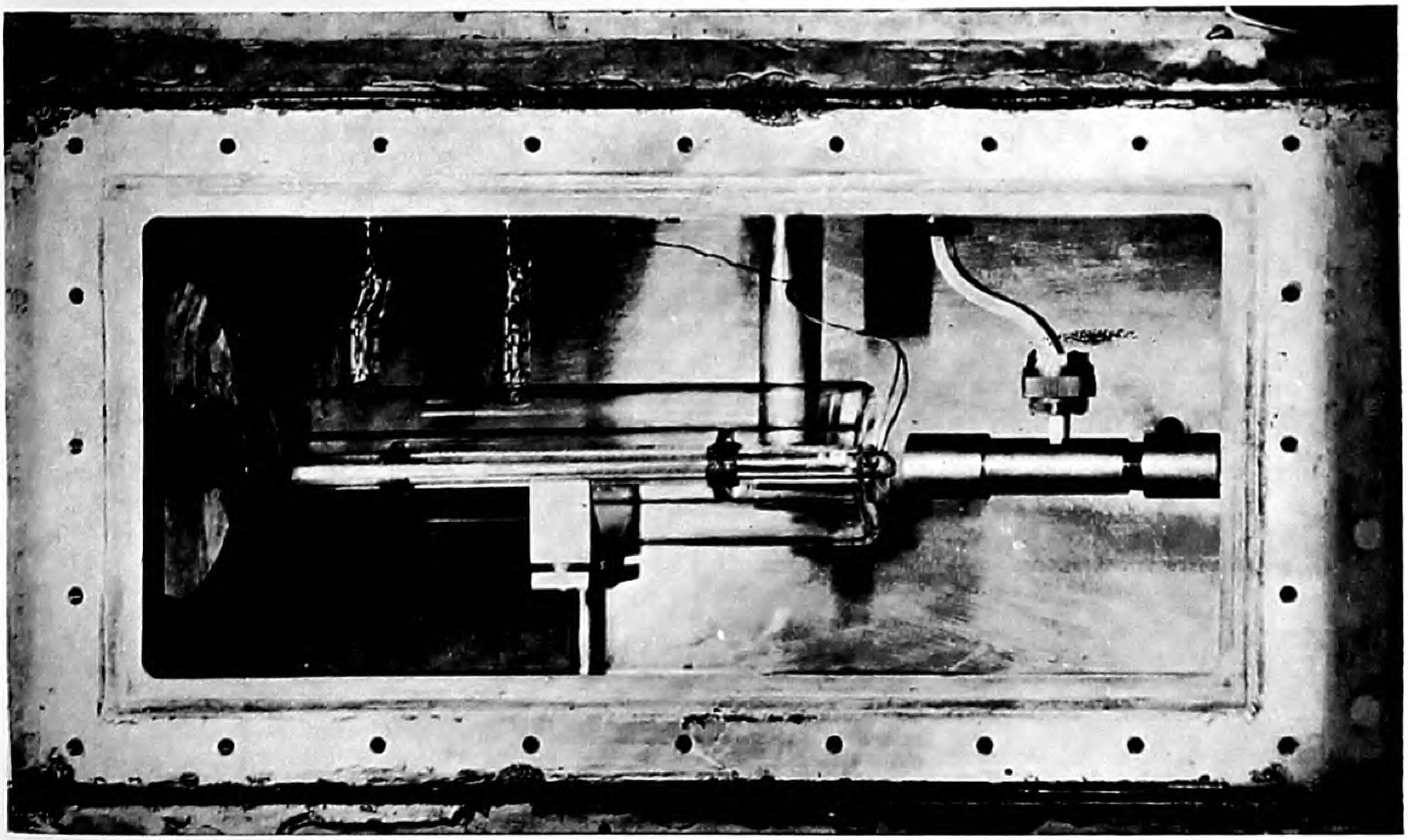
MASER PRINCIPLE is diagrammed above. Ammonia gas emerging from source (a high-pressure tank) contains high-energy molecules (*colored dots*) which emit radiation and low-energy molecules (*black dots*) which absorb it. The focuser is a ring of long electrodes. The electric field within the ring acts to disperse the low-energy molecules and concentrate the emitters. These find their way into the cavity, which is tuned to the frequency of the molecule. The oscillating energy is taken out through a wave guide.

tories, hopes to build a maser with radiating cesium atoms, and he estimates that this clock would be stable to one part in 10,000 billion for long periods.

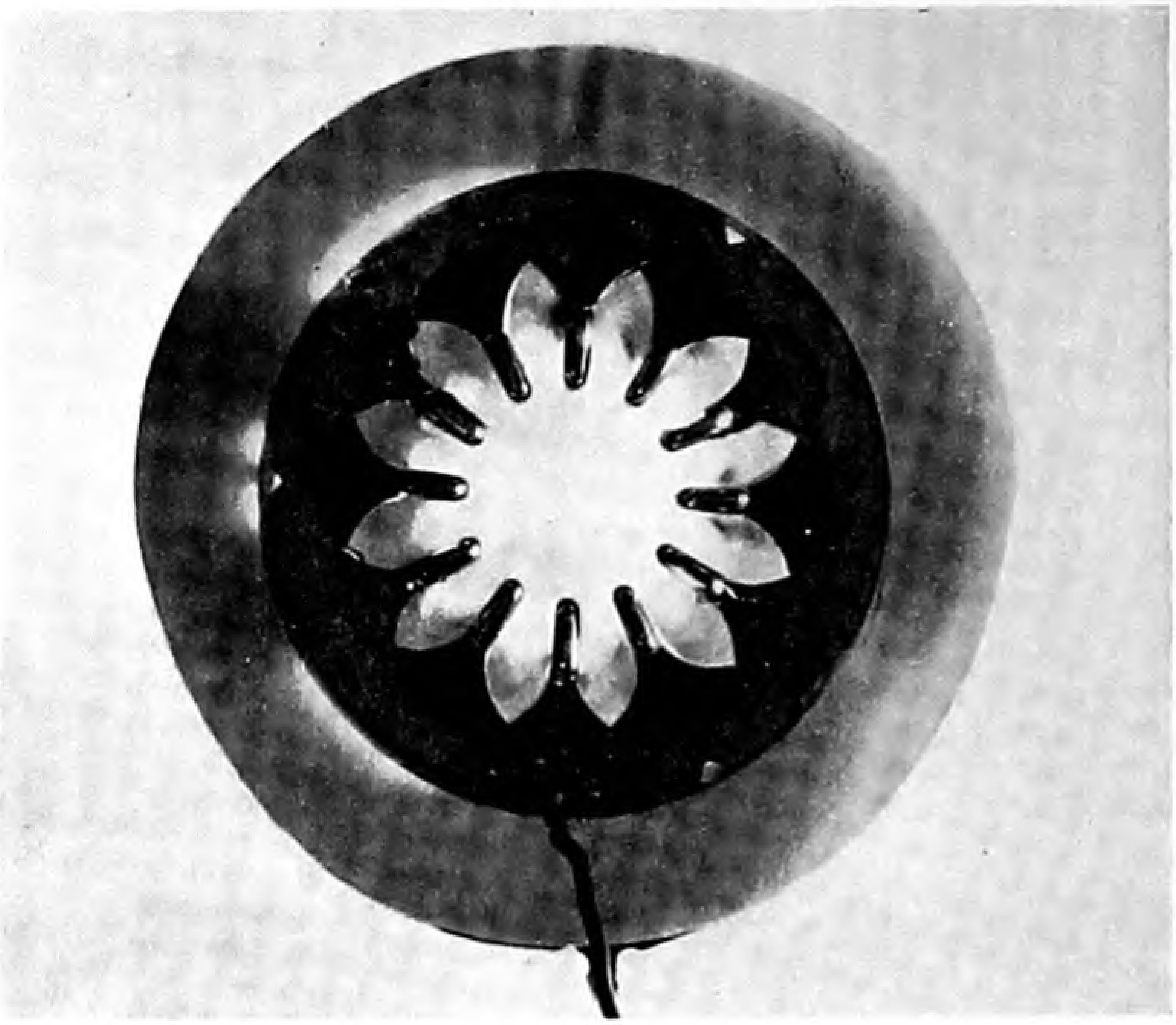
Uses for Atomic Clocks

When these remarkable timepieces have been built, what will they be used for? The list of needs is long and varied.

To begin with, atomic clocks would establish a more precise and invariant standard for the length of time units (*e.g.*, the second) than the astronomical one. The right time could be checked instantaneously, without waiting days or years for correcting astronomical measurements. The standard for distance could be related to the standard for time by means of an atomic clock coupled to an interferometer using microwaves. This would give the system of scientific units greater coherence and logic, for



ammonia "gun" is at left. Its gas stream is injected into the focuser, which is the assembly of horizontal rods in the center of the picture. The pipe at right is the resonant cavity, from which oscillating energy is withdrawn through rectangular pipe at upper right.



FOCUSER OF THE MASER, explained in the diagram on page 210, is photographed from the end. This component is from an early model of the maser built at Columbia University.

length and time are now measured in independent ways.

The establishment of a really accurate terrestrial time scale will permit more precise measurements of the earth's rotation, which in turn will help geophysicists to chart motions of the earth's molten interior, believed to be responsible for some of the irregularities in rotation. Atomic clocks also will play a major role in basic atomic research, making possible easier and more accurate measurement of the vibrations and rotations of molecules, atoms and nuclei.

Another imminent application is to aircraft navigation. Some of the present radio navigation instruments could give accurate position fixes over at least 3,000 miles if the frequency of the radio signals could be held stable to one part in a billion. Only 30 stations would be required to cover the entire globe.

The maser would even be useful in astronomy and cosmology. As a noiseless amplifier it would eliminate the noise generated in the circuits of radio telescopes and thus permit the resolution of weak signals, extending our vistas into space. Further, atomic clocks should make possible a check of the question whether the world of the atom runs on the same time as the universe.

There is a possibility that atomic clocks could furnish a test of Albert Einstein's general theory of relativity. The theory predicts that a light (or radio) wave traveling away from the earth should be slowed, or reduced in frequency, because of the work it does against gravity. A pair of atomic clocks, one at the bottom and the other at the top of a mountain, should be able to settle the point. The experiment would be of enormous interest, because there are few ways to check relativity theory.

Thus the atomic clock may reveal stories yet untold about our universe. Time will tell.

The Author

HAROLD LYONS is a leader in atomic research at the laboratories of the Hughes Aircraft Company. He was born in Buffalo in 1913. After taking a Ph.D. in physics from the University of Michigan, he went to work for the Naval Research Laboratories and later for the U. S. National Bureau of Standards, where he was chief physicist from 1946 to 1955. Lyons developed the first two atomic clocks, in response to the Bureau's need for an invariant measure of the time and frequency of microwaves. Lyons has won many honors for his work with atomic clocks. Among them are the U. S. Department of Commerce Gold Medal for Exceptional Service, the Arthur S. Flemming Award and an award from the Washington Academy of Sciences.

Bibliography

THE ACCURATE MEASUREMENT OF TIME. Dirk Brouwer in *Physics Today*, Vol. 4, No. 8, pages 6-15; August, 1951.

THE MASER—NEW TYPE OF MICROWAVE AMPLIFIER, FREQUENCY STANDARD, AND SPECTROMETER. J. P. Gordon, H. J. Zeiger and C. H. Townes in *The Physical Review*, Vol. 99, No. 4, pages 1,264-1,274; August 15, 1955.

MICROWAVE SPECTROSCOPY. C. H. Townes and A. L. Schawlow. McGraw-Hill Book Company, Inc., 1955.

A STUDY OF THE CHANGES IN THE RATE OF ROTATION OF THE EARTH. Dirk Brouwer in *The Astronomical Journal*, Vol. 57, No. 5, pages 125-146; September, 1952.

TIME AND ITS MEASUREMENT. G. M. Clemence in *American Scientist*, Vol. 40, No. 2; April, 1952.

SCIENTIFIC AMERICAN January 1957, Vol. 196, No. 1, pp. 84-92 OFFPRINT 226

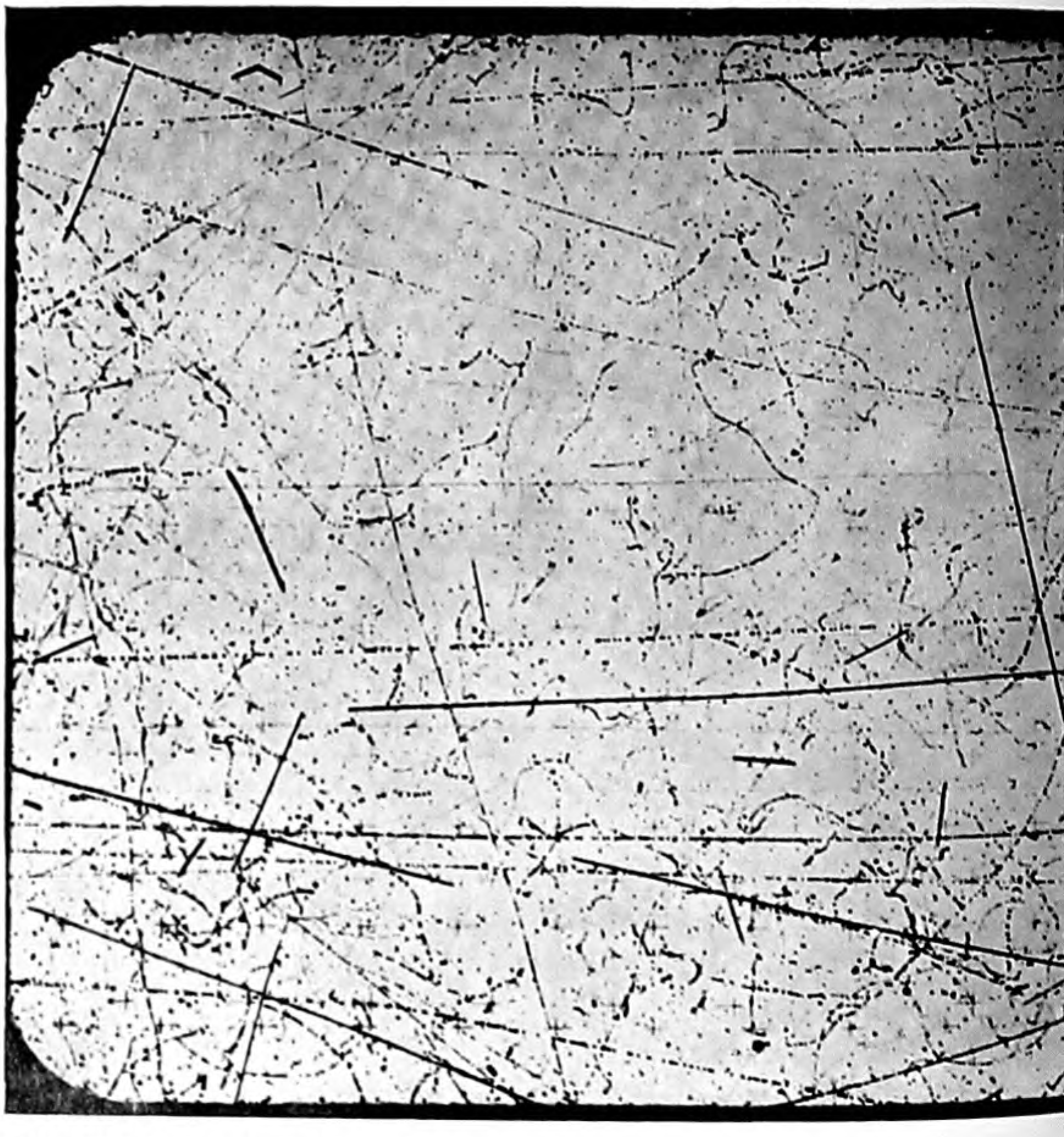
PIONS

by Robert E. Marshak

These particles, also known as pi mesons, are an important atomic structural material. They appear to be the cement which holds protons and neutrons together in the nucleus.

The cement that holds the universe together is the force of gravity. The glue holding the atom together is electromagnetic attraction. But the glue that holds the nucleus of the atom together is a mystery that defies all our experience and knowledge of the physical world. It is a force so unlike any we know that we can hardly find words to describe it. We do have a clue, however, to which we can give a name. It is the pi meson, or pion. In some way, not yet understood, pions are certainly involved in the nuclear binding force. Now that these particles can be generated at will by high-energy bombardment of matter, their properties are being industriously explored. What have we learned about them?

Before entering this strange realm, let us retrace very quickly the steps by which the physicists got there, along the now familiar path of electromagnetic forces. Considering the operation of these forces in the macroscopic world—as electricity, magnetism, light and so on—Michael Faraday and James Clerk Maxwell developed the concept of fields of force, pervading all space. When physicists began to examine the microscopic world of the atom, they assumed that the same field concept applied there as well: that the force between one electron and another, or between an electron and the positive nucleus, obeyed the laws of the classical electromagnetic field. But eventually it became clear that much of the behavior of atoms and electrons could be explained only on the assumption that the field in the atom is quantized. In other words, in the light of the quantum theory physicists concluded that electromagnetic forces are exerted by an exchange of quanta or packets of energy between charged bodies. These quanta are photons—massless units of energy. The field around an electron, say, consists of photons which the electron is continually emitting and absorbing. When one electron repels another, photons are interchanged; the photons are emitted by one particle and absorbed by the other. Thus the quantum theory, which is often said to do away with physical models, actually gives a more concrete picture of electromagnetic interaction than the classical theory did. Two charged bodies influence each other not through an intangible field but by tossing little pellets back and forth.



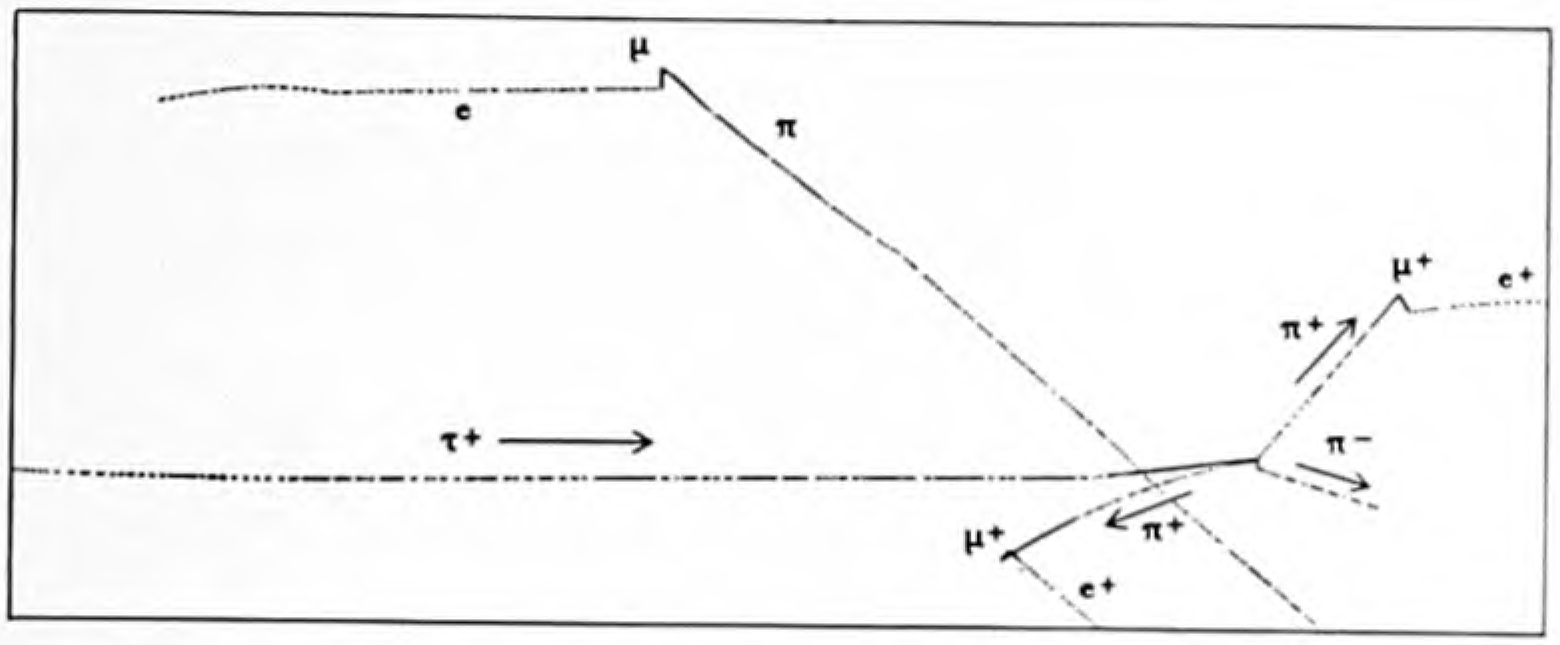
PION TRACKS are made visible as strings of tiny bubbles in a chamber of liquid propane. This photograph was made while the chamber was exposed to a beam of heavy mesons from the Brookhaven Cosmotron. As indicated in the drawing, one series of tracks shows

This conception was so successful in accounting for the forces on the atomic scale that the Japanese physicist Hideki Yukawa adopted it to attack the problem of the mysterious forces in the nucleus. He assumed that the force field in the nucleus, like that in the outer atom, is quantized. The attraction holding together protons and neutrons (nucleons) would thus be accounted for by a continual exchange of quanta of energy. But whereas the quantum of electromagnetic energy (the photon) is massless, Yukawa found that to explain the force of attraction in the nucleus, particularly its very short range, it had to be supposed that the nuclear quantum had an appreciable mass, which he calculated to be between 200 and 300 times that of an electron. To account for the great strength of the nuclear force, he assumed that the quanta were exchanged at a very rapid rate.

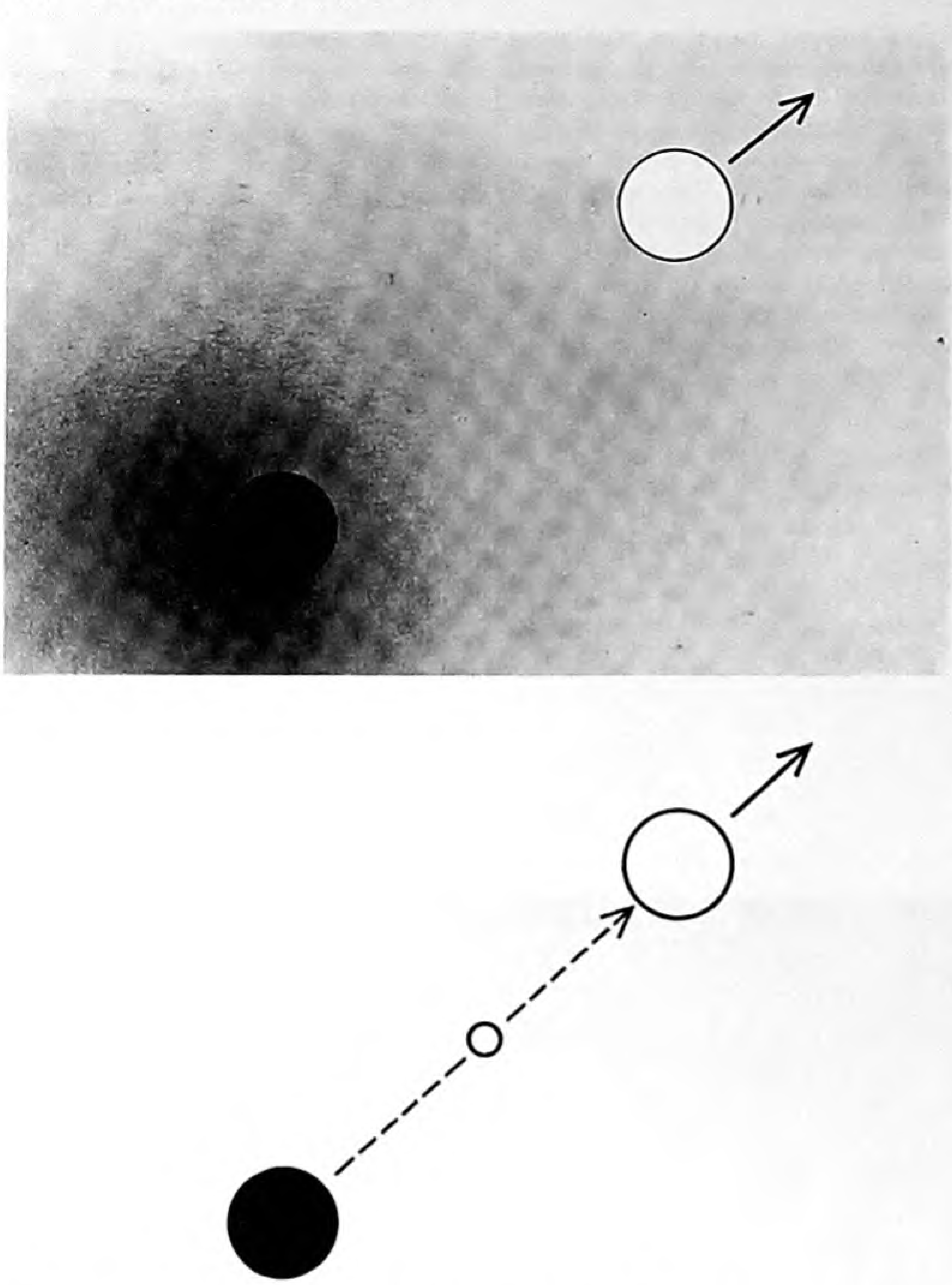
Yukawa's brilliant intuition was rewarded 12 years later (in 1947) by the discovery of the quanta, or particles, that he had proposed [see "The Multiplicity of Particles," by Robert E. Marshak; SCIENTIFIC AMERICAN, January, 1952]. The pi meson, or pion, has just the properties he predicted for it. Its mass is about 270 times that of the electron.

The notion that pions are exchanged between nucleons immediately raises some basic questions. To begin with, if a nucleon continually emits pions, what happens to the conservation of mass? The emission of a pion, with its appreciable mass, should reduce the mass of the nucleon, and yet in all our experiments the mass of a nucleon remains constant. The answer is that the emission and reabsorption of pions takes place so rapidly that we cannot detect it. Since any phenomenon that is undetectable cannot be regarded as "real," in the physical sense, we must speak of "virtual" emission and exchange of pions.

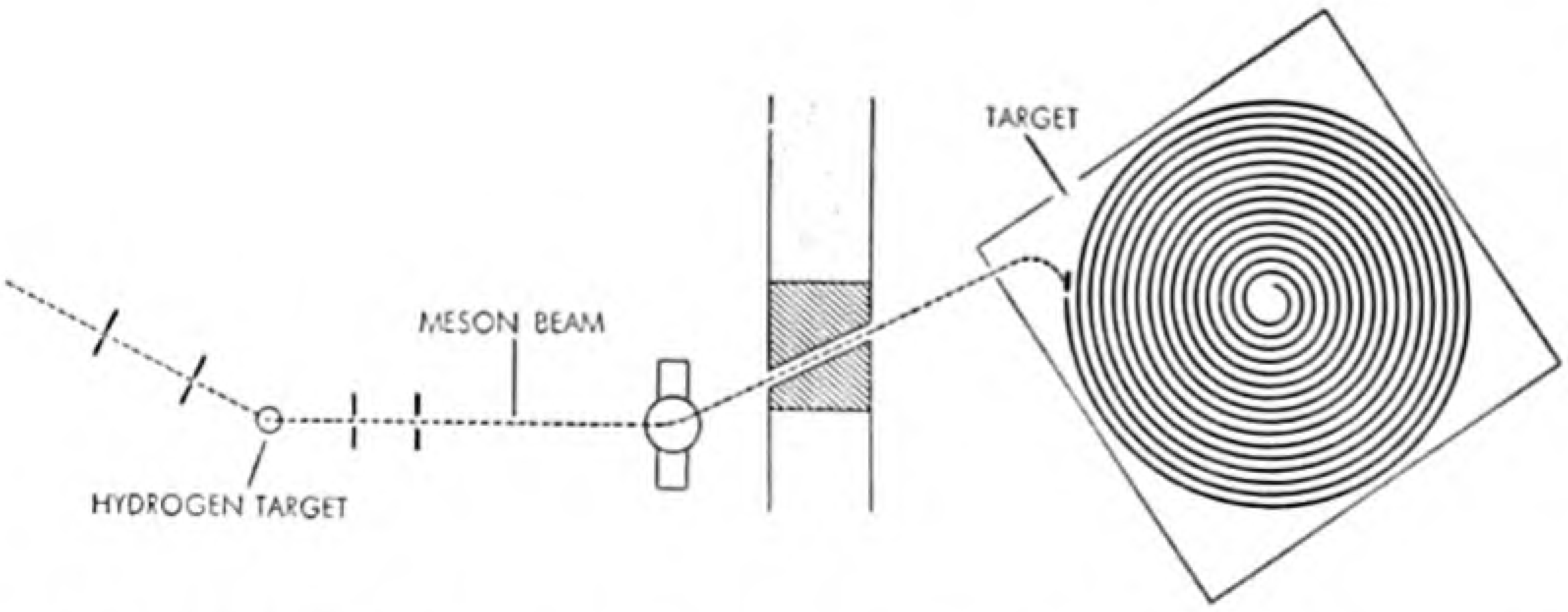
To understand a little more clearly what this means, and to see approxi-



a tau meson (τ^+) entering from the left and decaying into two positive pions (π^+) and a negative pion (π^-). The positive pion decays into a mu meson (μ^+), which decays into an electron (e^+). Another pion enters the chamber from the lower right and decays into a mu meson at upper left. The experiment was performed by D. A. Glaser and his colleagues of the University of Michigan.



ELECTROMAGNETIC FIELD around a charged particle such as an electron (*black circle*) is shown schematically according to classical theory (*above*) and quantum theory (*below*). The gray shading in the upper diagram represents the continuous classical field through which the particle was thought to exert its force on a second electron (*white circle*). The circle in lower diagram shows the field quantum or photon now thought to transfer the force.



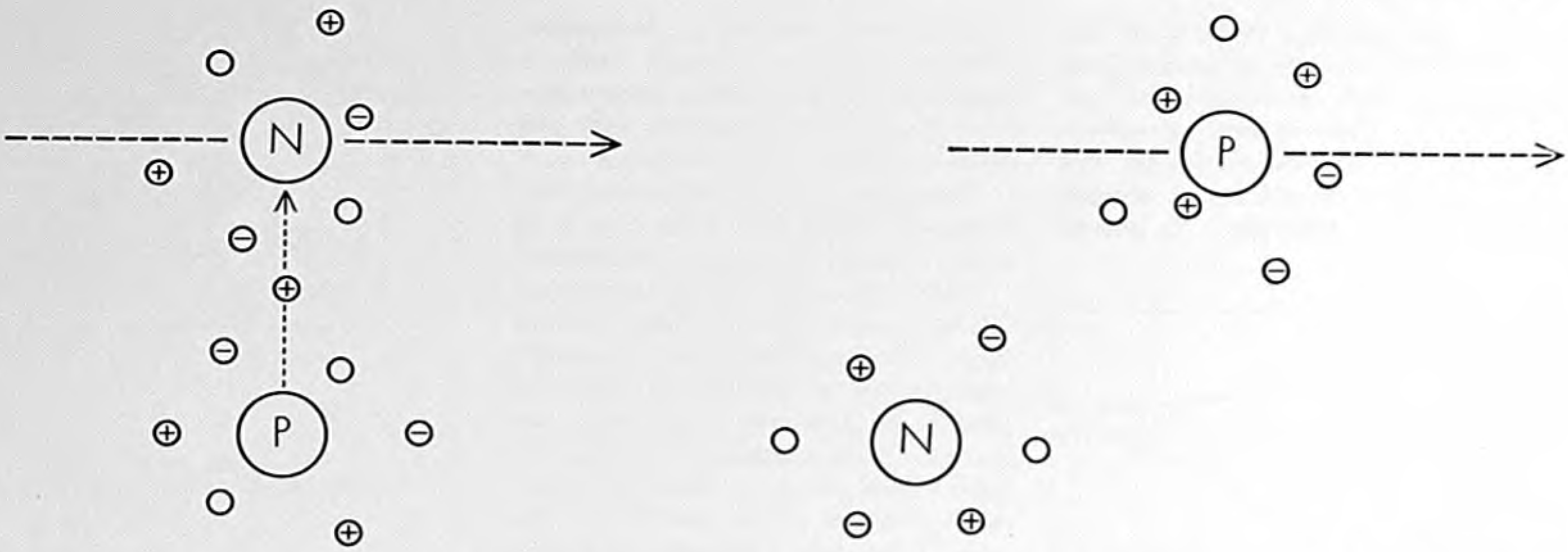
PION SCATTERING experiment is diagrammed above. The meson beam from a cyclotron (*top*) is steered by a magnet to strike a sample of hydrogen. The counters in front and back of the target record the numbers of particles that are deflected at various angles.

mately how brief the appearance of a pion must be, we must recall the famous uncertainty principle. This principle sets a definite limit to our knowledge of very small-scale phenomena. It says, for example, that if we measure an electron's *position* exactly, we thereby destroy our ability to make any measurement whatever of its *momentum* and *vice versa*. If we want figures for position and momentum at the same time, we must settle for inexact measurements of both. The uncertainty principle of quantum theory tells us the maximum accuracy we can hope to attain: the uncertainty in position multiplied by the uncertainty in momentum must be at least as great as the value of Planck's constant, h.

Another pair of quantities which fall under the principle are energy and time. Any experiment for measuring the energy of a system requires a certain time to perform. Any such experiment also tends to alter the energy. Now it turns out that the shorter the time of measurement, the greater the effect on the energy. In other words, the more certain we are about the time at which the energy is determined, the less certain we can be about its amount. Again, the product of the two uncertainties can never be less than h.

The energy equivalent of the mass of a pion is 135 million electron volts. The calculation based on h tells us that the margin of uncertainty for the energy content of a nucleon will be at least 135 Mev when the time of measurement is 5×10^{-24} of a second. Hence if a pion is emitted and reabsorbed within this time, it will be undetectable, *i.e.*, "virtual." In that time a pion, moving at practically the speed of light, could travel to about 1.4×10^{-13} of a centimeter from the center of a nucleon. This distance is just about the observed range of nuclear forces! The agreement seems a striking support for the theory that pions act as agents of these forces.

This conception of the pion accounts for another phenomenon observed in experiments. When hydrogen is bombarded by a beam of fast neutrons, many protons (hydrogen nuclei) shoot out in the forward direction at about the speed of the impinging neutrons, while a corresponding number of neutrons is found almost stationary within the target. It is altogether improbable that so many neutrons would hit protons dead center and transfer all their momentum to the protons. A much more plausible interpretation is that the emerging protons represent neutrons which were converted into protons during passage through the tar-



NUCLEAR FORCE FIELD consists of clouds of virtual pions (*small open circles*) which surround each nucleon. At left a moving neutron (*above*) passes close to a stationary proton (*below*) and picks up a positive pion. The moving particle is now a proton and the stationary one a neutron. By transferring neutral pions, nucleons are also able to interact without this exchange of charge.

get. Such a conversion could occur if a neutron seized a positively charged pion: the neutron would thus become a proton, and the proton that lost the positive charge would become a neutron. Thus a neutron dashing through a proton target grabs a pion from a proton and leaves a neutron behind.

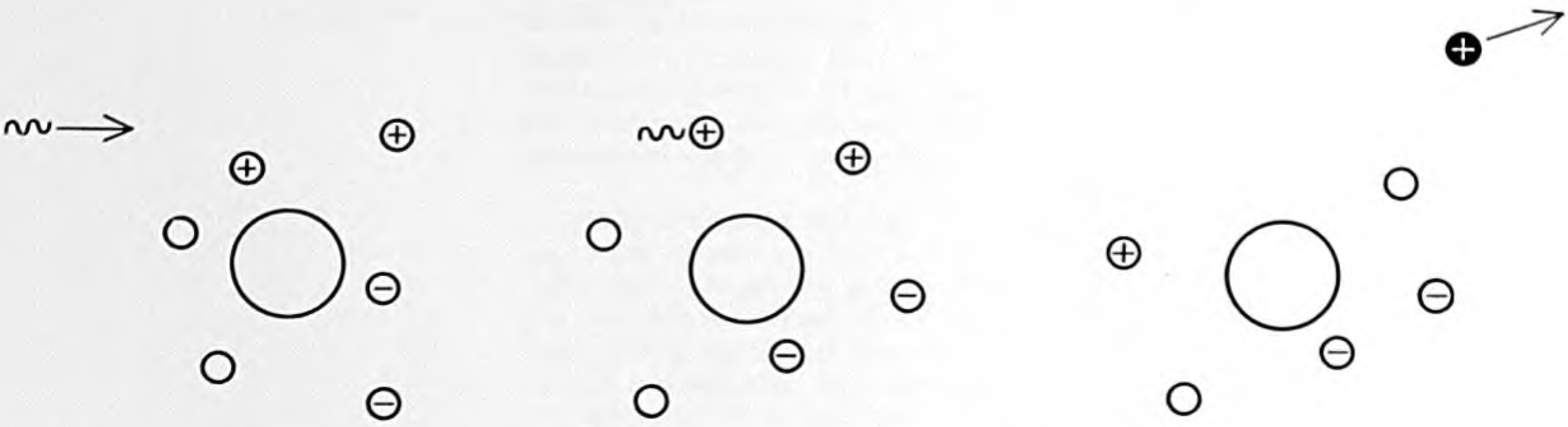
The attraction between nucleons in the nucleus of an atom (other than ordinary hydrogen, which has only one nucleon) may be exerted through such exchanges of pions. There should be positive, negative and neutral pions, to bind together protons with protons, protons with neutrons and neutrons with neutrons.

How is it that the ghostly "virtual" pions are in fact detectable as real particles? If we consider pions as packets of energy, it is not difficult to describe the circumstances under which we should be able to detect them. Suppose the energy equivalent of a pion is supplied to a nucleon, replacing the pion energy. A pion may then be released and detected before it is captured by another nucleon. Pions were first identified in the debris from cosmic ray collisions and then manufactured in high-energy accelerators by bombardment of nuclei.

All three forms of the pions have been found—positive, negative and neutral. The neutral pion has 264 times the mass of the electron. The charged pion, which gains a little mass from its interaction with the electromagnetic field, has a mass of 273. Pions are readily absorbed by nuclei. They are unstable. The charged pion decays into a lighter particle (called the mu meson) and a neutrino with a half-life of a few hundred-millionths of a second. The neutral pion decays much faster (half-life about 10^{-15} of a second) into two gamma rays.

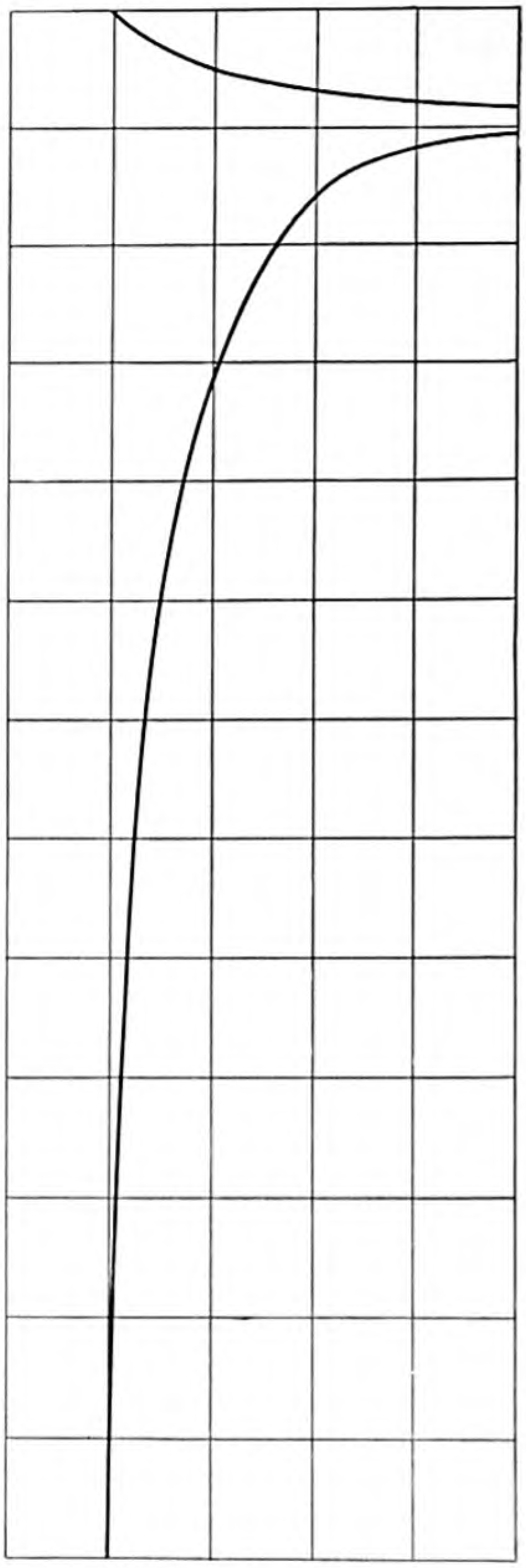
Pions, produced in large quantities by bombarding targets such as carbon, are now formed into beams for probing nuclei. The way in which a beam of real pions is deflected or scattered by the target nuclei shows how the pions interact with the nuclear force field. The simplest and most informative experiments are those on individual nucleons. Protons can be studied directly by bombarding a target of hydrogen. Neutrons are examined by bombarding heavy hydrogen, or deuterium, whose nucleus contains one proton and one neutron. When the proton's effect is subtracted out, the remaining pattern gives the neutron's interaction with the pion.

These experiments support a picture which, if rather obscure, at least fits together well as far as it goes. For instance, investigators of the nuclear forces have been interested in the fact that the force of attraction between nucleons seems to be exactly the same whether they are charged or uncharged (*i.e.*, protons or neutrons). This of course is entirely for-



VIRTUAL PION MATERIALIZES if the system is provided with an amount of energy at least equivalent to the pion's rest mass. In these diagrams a high-energy gamma ray (*wavy line*) enters a meson cloud and strikes a virtual positive pion (*open circle*). The energy of the gamma ray is absorbed and the virtual pion turns into a real pion (*solid circle*). It can now escape from the nucleon.

eign to our experience in the world outside the atomic nucleus. Considering the hypothetical roles of charged and neutral pions in these several attractions, the investigators have calculated that nucleons must emit and absorb charged pions twice as frequently as neutral pions. To put it another way, the nucleon interacts twice as strongly with a charged pion field as with a neutral pion field. Scattering experiments with real pions have confirmed this assumption.



FIELD STRENGTHS due to electromagnetic and nuclear forces are indicated by these curves. Upper curve represents the electromagnetic field around a proton (which repels another proton). Lower curve is proton's nuclear field (which attracts another proton). Horizontal axis gives distance from proton in units of 1.4×10^{-13} centimeters. Vertical scale is arbitrary.

The nuclear force of attraction is very strong—so strong that, if the pion is its agent, a nucleon must emit virtual pions at a high rate, and must be surrounded at close quarters by a veritable cloud of them. This suggests that if enough energy could be supplied, it should be possible to materialize more than one real pion from a nucleon. This has indeed turned out to be true. The three-Bev Cosmotron at the Brookhaven National Laboratory produces an average of two pions per collision. The six-Bev Bevatron at Berkeley gives an average of more than three. In primary cosmic radiations, where there are energies as high as 10,000 Bev, as many as 20 real pions have been observed to emerge from a single collision.

Thus the proton and the neutron, once supposed to be the ultimate building blocks of matter, become less monolithic than they seemed. Each consists of a core surrounded by a fluctuating cloud of pions—an arrangement somewhat reminiscent of the atom with its nucleus and planetary electrons. Pion scattering experiments now indicate that nucleons may even have excited states, as the atom does. The excitation, produced when a pion hits a nucleon violently, takes the form of a temporary increase in the nucleon's charge. Presumably this state involves some rearrangement of the meson cloud, but the details are not yet known.

The inner region near the core of the nucleon, where the meson cloud is most dense, remains an area of mystery. We can probe it with very fast particles: the faster the projectiles, the deeper they penetrate. But our knowledge and theories about the meson field are still too vague to yield equations which might account for the scattering patterns observed or predict how many pions will be produced at a given bombarding energy.

The chief difficulty is the fact that we must deal with a swarm of pions. Our mathematical techniques cannot effectively handle more than one pion at a time. Beyond this things become much too complicated. The problem appears to be a basic one, and it seems that only some radically new idea will enable us to solve it. And so the pion, while providing a tantalizing glimpse into the nuclear forces, serves also to deepen our ignorance.

The Author

ROBERT E. MARSHAK is Harris Professor of Physics and chairman of his department at the University of Rochester. He is intimately associated with the discovery of the pi meson; he predicted its existence in his two-meson theory, which he set forth in 1947 when only the mu meson had been found. Marshak was born in New York City and attended Columbia University as a Pulitzer Scholar, graduating at the age of 19. At 22 he took a Ph.D. in theoretical physics under Hans Bethe at Cornell University and joined Victor F. Weisskopf at Rochester. During the war he worked for the Manhattan District. Marshak edits *Interscience Tracts in Physics and Astronomy* and has written a book on meson physics and contributed articles to SCIENTIFIC AMERICAN.

Bibliography

ELEMENTARY PARTICLES. Enrico Fermi. Yale University Press, 1951.

THE LAWS OF NATURE. R. E. Peierls. George Allen & Unwin Ltd., 1955.

MESON MASSES AND ENERGETICS OF MESON DECAY. Walter H. Barkas in *American Journal of Physics*, Vol. 20, No. 1, pages 5-13; January, 1952.

MESON PHYSICS. R. E. Marshak in *Annual Review of Nuclear Science*, Vol. 1, pages 1-42; 1952.

MESONS AND FIELDS: VOLUME TWO. Hans A. Bethe and Frederic de Hoffmann. Row, Peterson and Company, 1955.

MESONS: A SUMMARY OF EXPERIMENTAL FACTS. Alan M. Thorndike. McGraw-Hill Book Company, Inc., 1952.

MR TOMPKINS EXPLORES THE ATOM. George Gamow. The Macmillan Company, 1944.

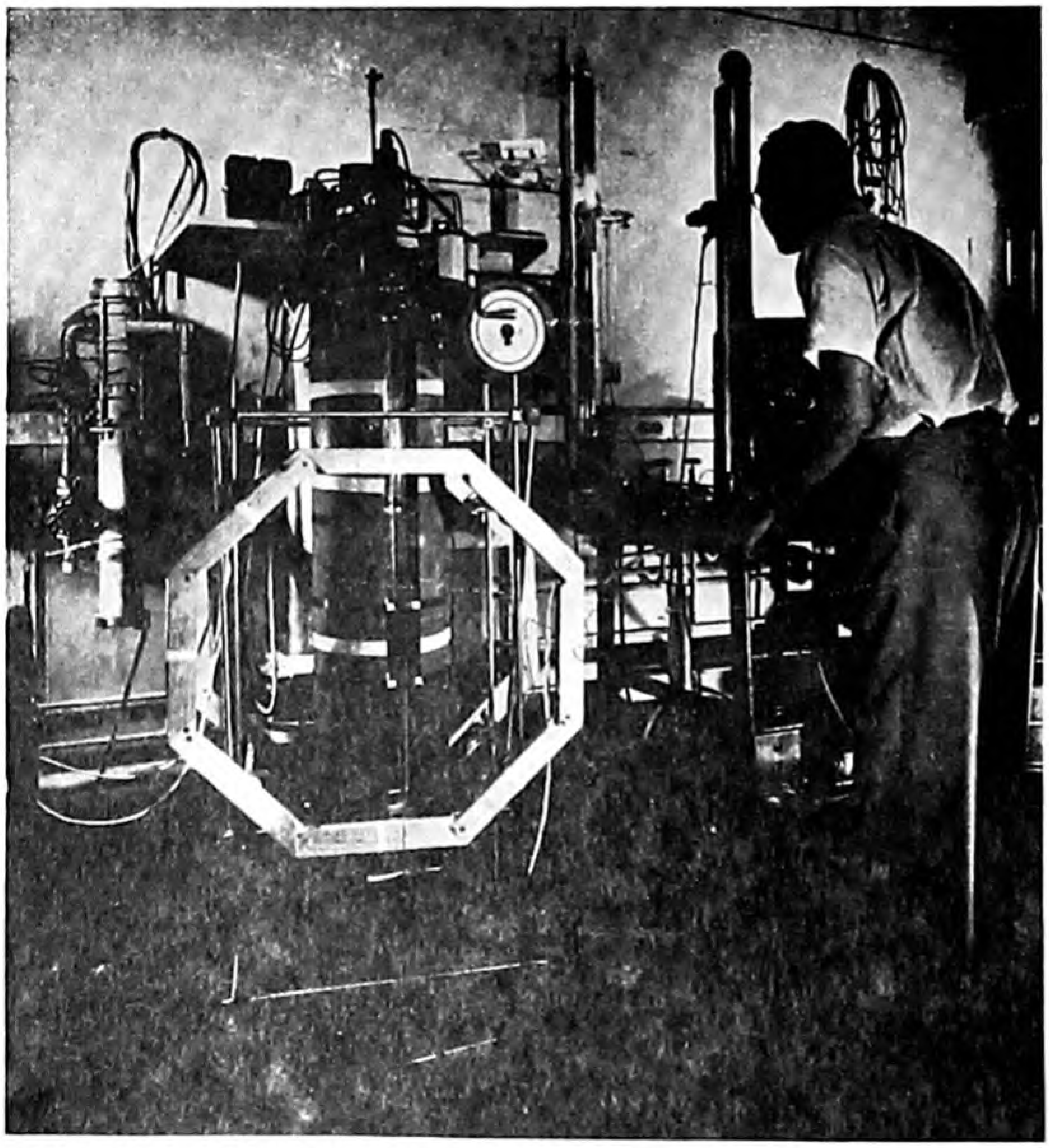
SCIENTIFIC AMERICAN November 1957, Vol. 197, No. 5, pp. 92-103 OFFPRINT 227

SUPERCONDUCTIVITY

by B. T. Matthias

When certain metals are cooled to a few degrees above absolute zero, they lose all electrical resistance. Although the phenomenon has not been explained, it has now been found to adhere to some simple rules.

It was almost 50 years ago that Heike Kamerlingh Onnes, a Dutch investigator of the behavior of matter at very low temperatures, passed an electric current through some frozen mercury and made a startling discovery. He found that, at a few degrees from absolute zero, all resistance to the flow of the current disappeared. Physicists today are still hunting for an explanation of this bizarre phenomenon of superconductivity, which seems to contradict some of our basic ideas about nature. But though we do not understand it very well, we are beginning to learn enough about superconductivity to make use of it. This article is a report of some recent work which suggests that it may be possible to make superconductors which can be used in electrical equipment held at low temperatures.

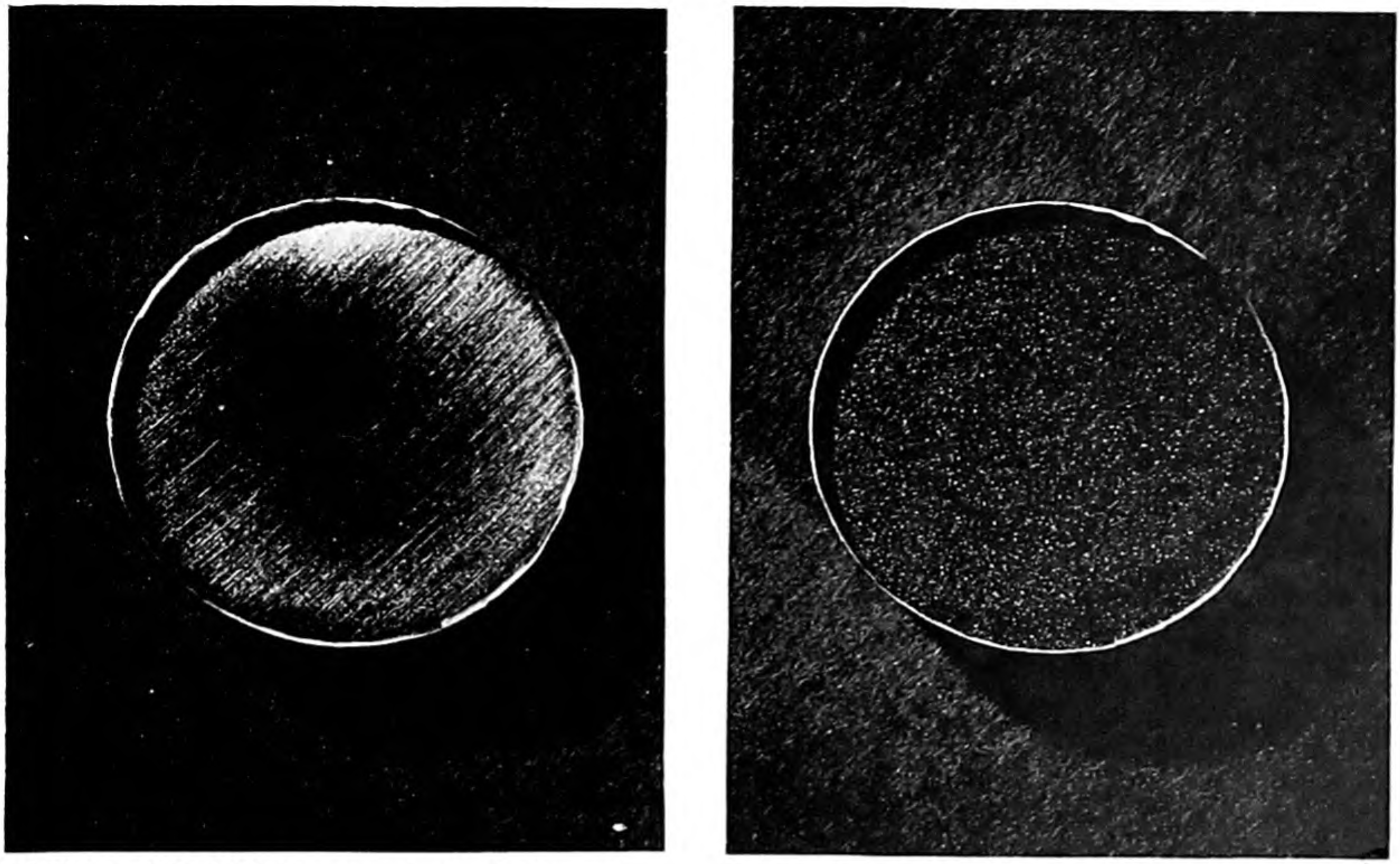


EFFECT OF A MAGNETIC FIELD on superconducting metals is studied with this apparatus in the laboratory of A. L. Schawlow at Bell Telephone Laboratories. A sample of superconducting metal is placed in liquid helium at the bottom of the cylinder at left.

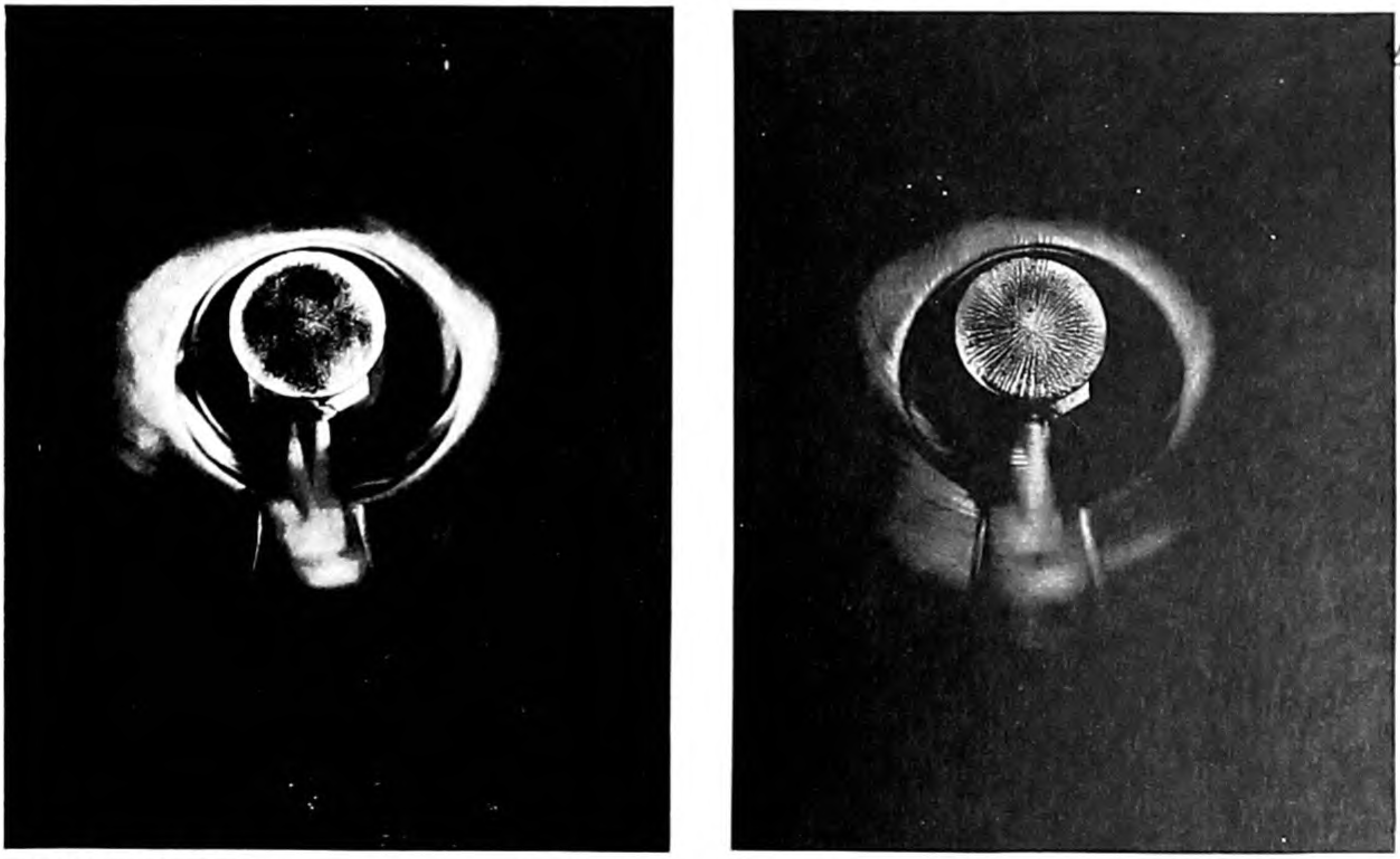
Let us begin by considering the normal conduction of an electric current by a metal. We know that the current is transmitted by the motion of electrons, driven through the metal's crystal lattices by the applied electric voltage. The electrons collide with the atoms in the lattice; this impedance of their motion constitutes the conductor's electrical resistance. The resistance increases as the temperature increases, because the vibrating atoms in the lattice then oscillate over wider distances from their lattice positions and interfere with the electrons' motion more strongly.

Now it is reasonable to suppose that if the atoms' vibrations were completely stilled by reducing the temperature to absolute zero, resistance to the flow of electrons might drop to an undetectable level. But Kamerlingh Onnes found that resistance vanished abruptly at several degrees above zero. Mercury became superconducting at 4.2 degrees Kelvin; certain other metals did so at different temperatures in the low range down to around one degree—the lowest Kamerlingh Onnes could reach. Later the German physicist W. Meissner discovered another peculiar property of metals in the superconducting state. They were impervious to a magnetic field: placed between the poles of a magnet, they completely expelled the field, so that the lines of magnetic force went around the material [*see diagram at top of page 222*].

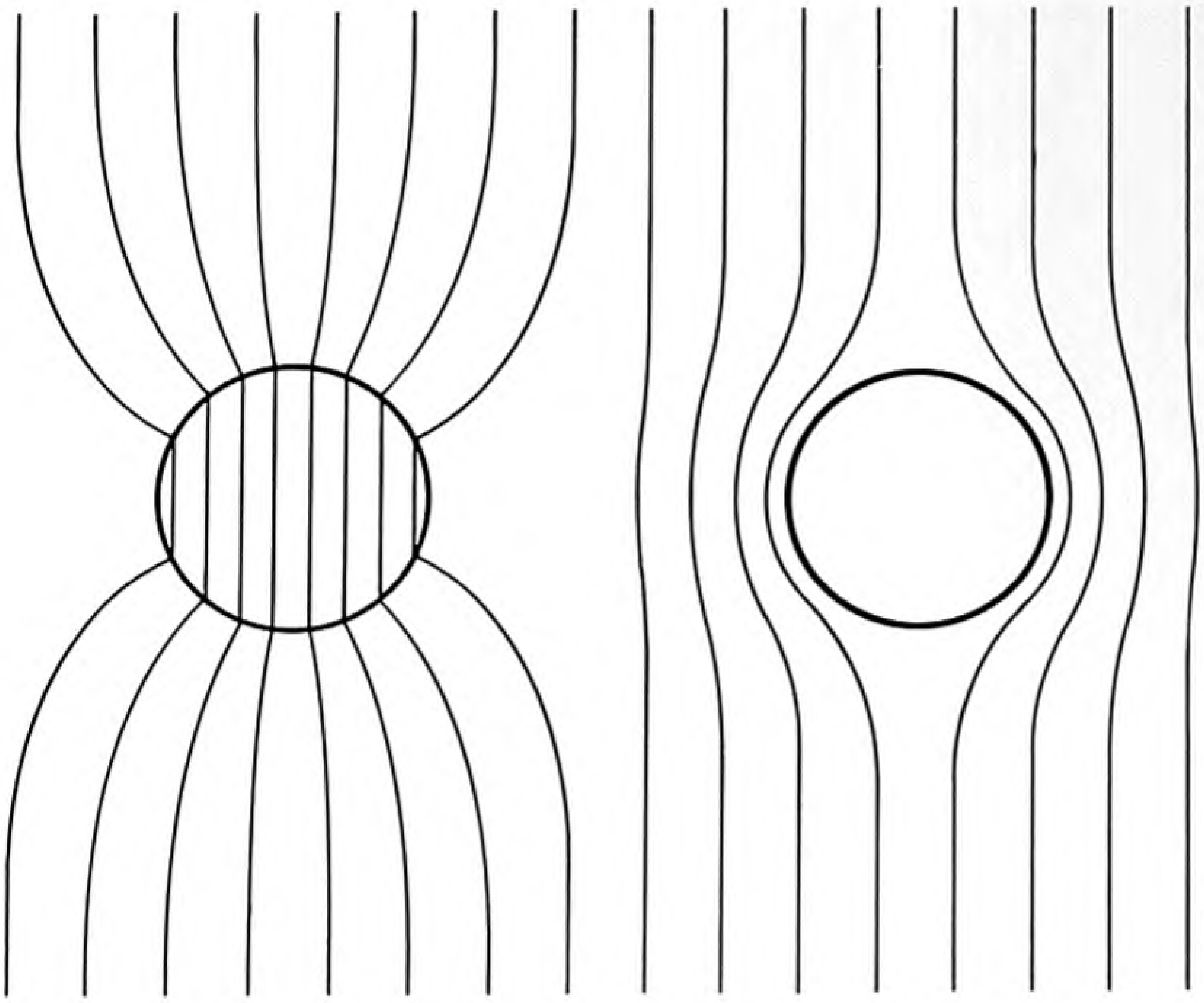
About 20 years ago Fritz London, a German-born physicist then working in England, suggested a theory



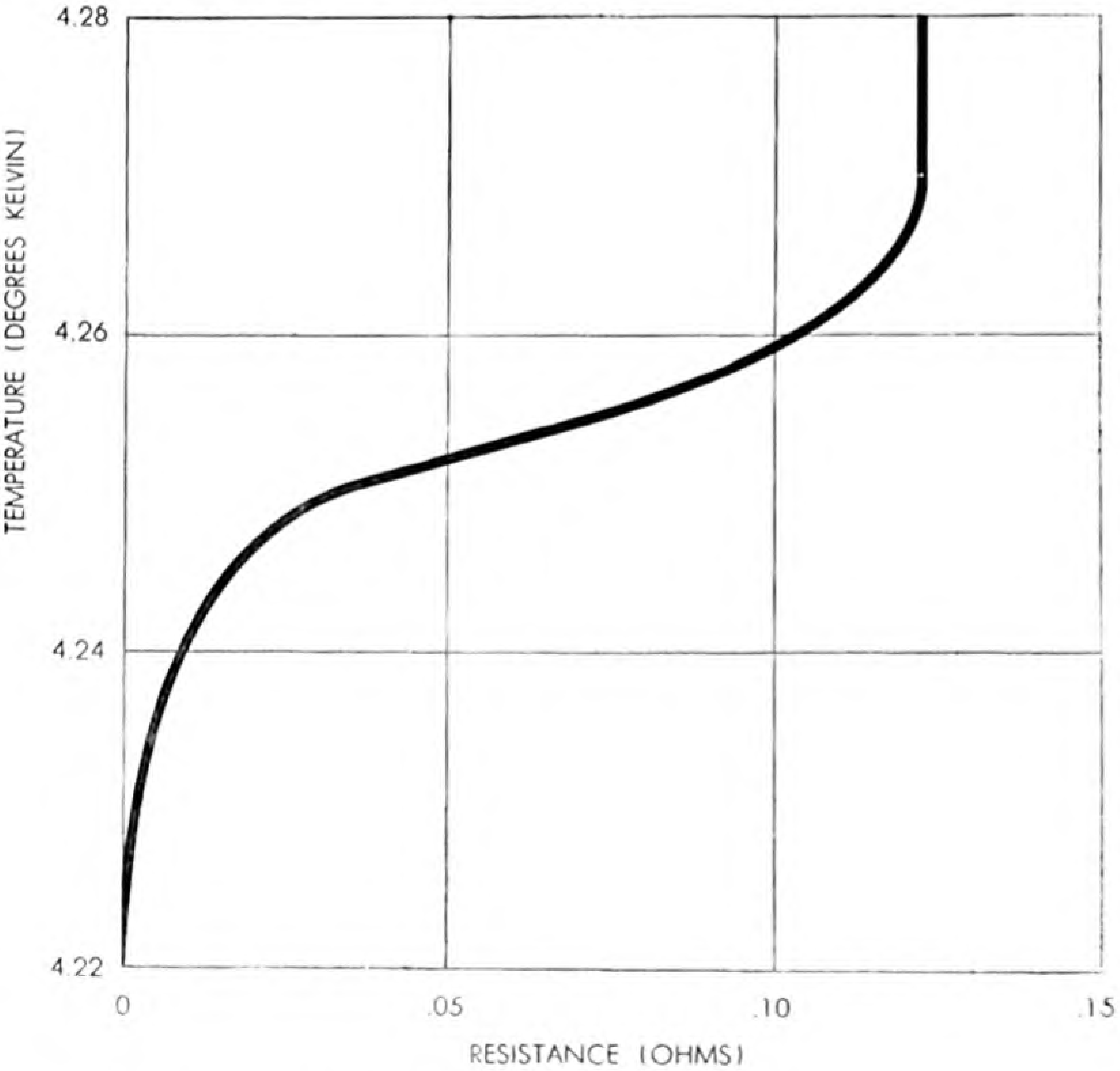
SAMPLE IS PREPARED for an experiment with the apparatus on the opposite page. At left is a disk of very pure tin surrounded by a low retaining wall made of paper. At right the surface of the disk has been covered with a powder of the element niobium.



SAMPLE IS LOWERED into apparatus (*left*). Around it is the silvered wall of an insulating Dewar flask. After an electromagnet around the flask has been switched on, the particles of niobium move into superconducting and non-superconducting areas (*right*).



EFFECT OF SUPERCONDUCTIVITY on a magnetic field is depicted in these schematic drawings. A sphere of ferromagnetic material (*circle at left*) concentrates the lines of force in a magnetic field. A sphere of superconducting material (*circle at right*) expels the field.



ELECTRICAL RESISTANCE OF MERCURY is plotted against temperature. The resistance disappears entirely at about 4.2 degrees Kelvin (4.2 degrees centigrade above absolute zero).

about the behavior of matter in the superconducting state. Since the free movement of electrons through a superconducting metal is analogous to the unimpeded motion of electrons in their orbits around the nucleus of an atom, London reasoned that an external magnetic field fails to penetrate a superconducting material because it shifts the large "orbits" of the electrons in the material so that they set up a counteracting magnetic field of their own.

London's concept gave an understandable picture, but of course it did not explain why a mass of material should behave like a single giant atom—that is to say, why reduction of the material to a very low temperature should create unobstructed orbits for the flow of electrons. In 1950 H. Fröhlich at Purdue University and John Bardeen of the Bell Telephone Laboratories attempted a more specific explanation. Their theory said that at low temperatures the vibrating atoms in a crystal lattice no longer obstruct the flow of electrons but on the contrary begin abruptly to conduct this flow in a wave-like way; in other words, the lattice vibration itself becomes the agent that makes the metal superconducting.

According to the Heisenberg uncertainty principle in the quantum theory, the vibrations of atoms can never die away entirely. Even at absolute zero the atoms in a lattice would retain an irreducible motion, called "zero-point vibration." The Fröhlich-Bardeen theory of superconductivity holds that at very low temperatures these vibrations of atoms and the motions of electrons are synchronized, so to speak. As a result of the interaction between the atoms and the electrons, the electrons reduce their energy and ride along with the lattice vibration as on a wave. Thus resistance to the flow of electrons disappears.

The theory was suggested by the fact that metals which are comparatively poor conductors at normal temperatures are most apt to be superconductors at low temperatures. Fröhlich and Bardeen reasoned that since these metals have a strong scattering effect on electrons at elevated temperatures, they should have a strong ordering effect on electrons when the vibrations of their atoms and the motions of electrons become coordinated at low temperatures. On this reasoning, materials whose vibrating atoms interact strongly with electrons should become superconductors more readily (*i.e.*, at higher temperatures above absolute zero) than those

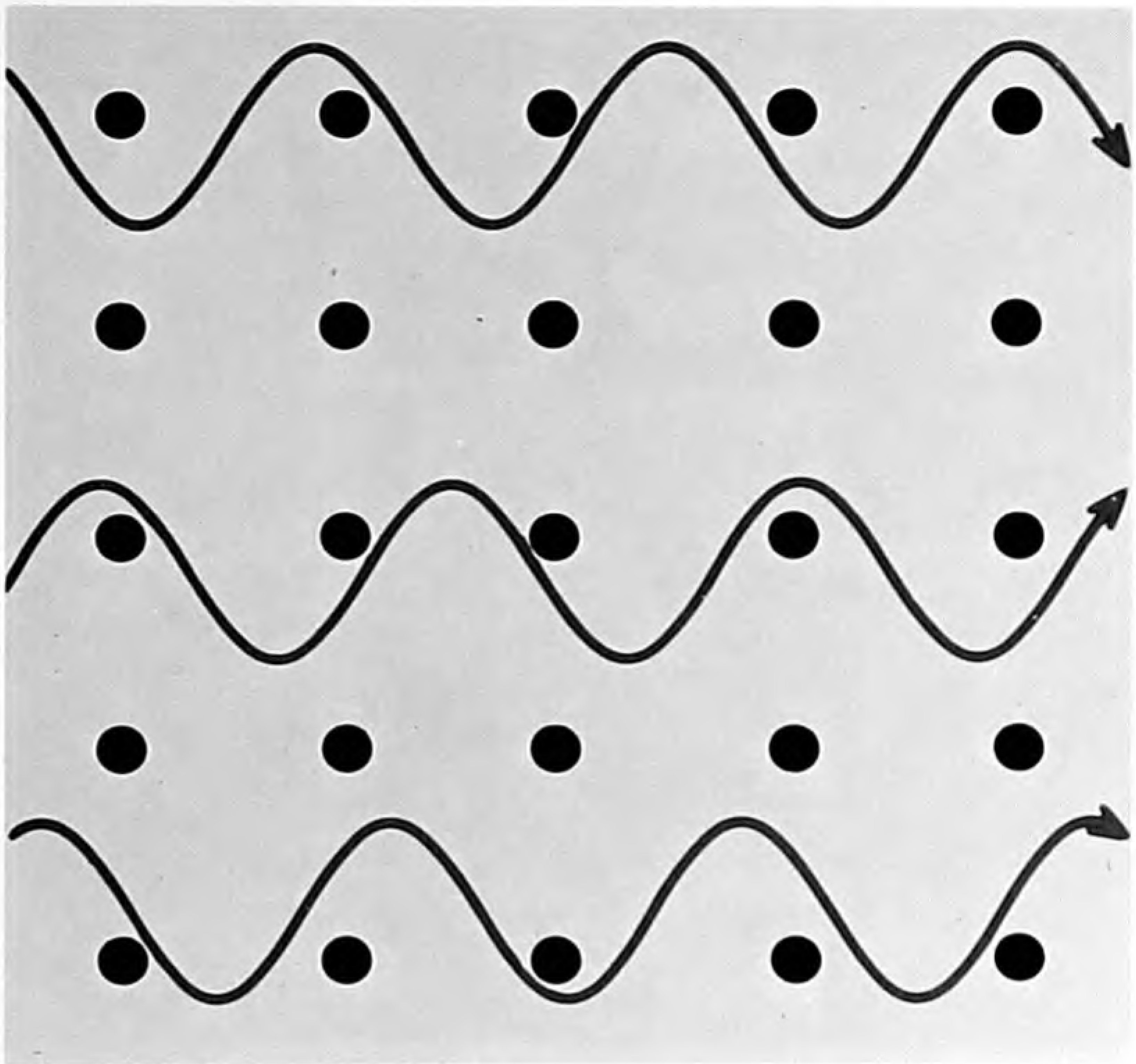
whose atoms interact weakly with electrons. As a corollary, the heavier the element, the less likely it is to become a superconductor, because its low-temperature vibrations will be comparatively slow. Fröhlich predicted that light isotopes of an element would become superconducting sooner (at higher temperatures) than the heavier ones, and this proved to be true.

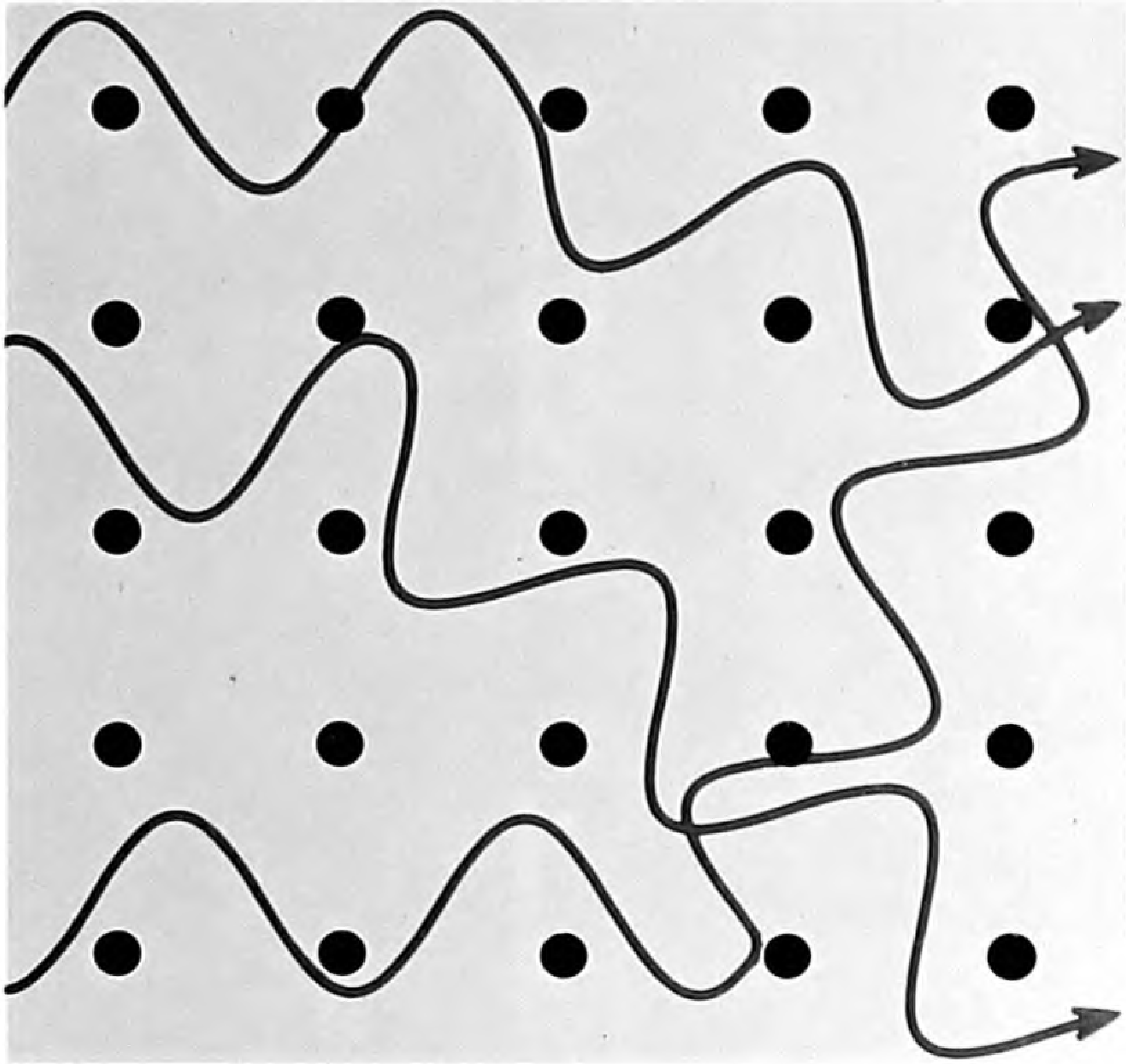
Recently Bardeen, Leon N. Cooper and J. Robert Schrieffer at the University of Illinois worked out a more complete new theory. It yields a correct description of properties of a superconductor, but there is still some question about the mathematical rigor of its treatment of the magnetic effect.

In 1950 John K. Hulm and I decided to try an entirely different approach. Since the theories offered no dependable means of predicting just what substances might be superconductors, or at what temperatures they would reach this state, we set out to find superconducting materials strictly by experiment. Our hope was that after we had tested a large number of substances, we might begin to see a pattern which would identify some of the physical and chemical properties associated with superconductivity.

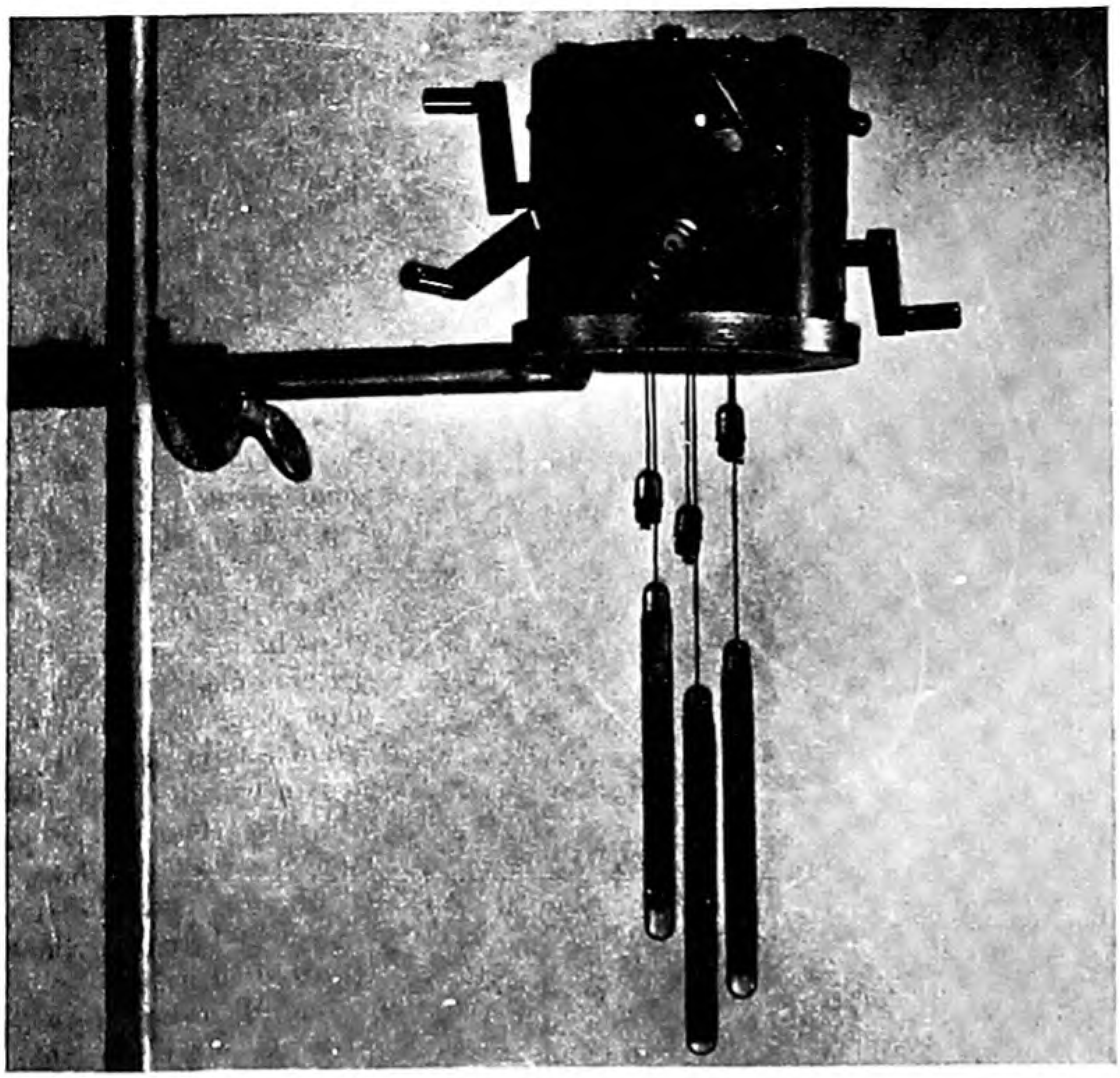
Having almost no clues as to where to look, we tried one compound after another. At first we seemed to be getting nowhere. But after three years of testing a great number of compounds a crucial fact began to emerge. It appeared that the decisive factor determining how readily a substance would become superconductive was the number of valence electrons possessed by its atoms—valence electrons being those in an outer unfilled shell. We found that the only substances which became superconducting were elements or compounds with an average of between two and eight valence electrons per atom. And within this range the materials with an odd number of valence electrons per atom—three, five or seven—become superconducting most easily (farthest above absolute zero).

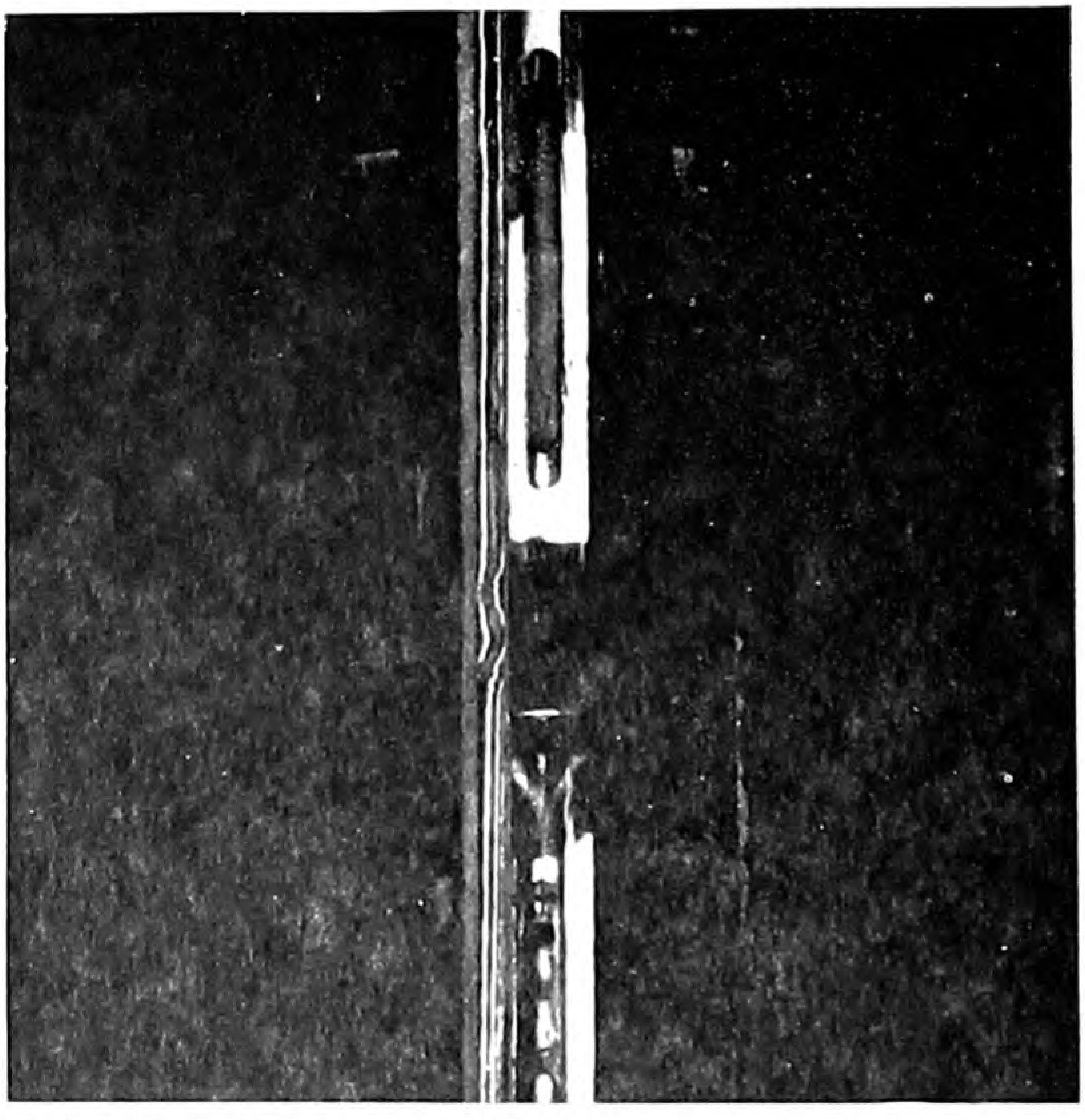
Here is the kind of rule we have been looking for. We now have a specific guide to finding or synthesizing superconductors. We should seek substances with an average of three, five or seven valence electrons per atom. To this rule we can add a few other helpful clues: it is known, for instance, that superconductivity is favored by certain kinds of crystal structure and by the amount of empty space in the crystal (*i.e.*, space





SUPERCONDUCTING ELECTRONS (*wavy lines at top*) interact in an orderly way with atoms (*dots*) in a crystal. Ordinary conducting electrons (*bottom*) are deflected by atoms.





SAMPLES IN CAPSULES at top are tested by Matthias for superconductivity by lowering the capsule into a coil cooled with liquid helium. In this photograph at the bottom a capsule may be seen through a window in the apparatus just before it is lowered into the coil.

not occupied by atoms).

With these empirical rules to steer us, we have been able to produce many superconducting materials. I shall illustrate with a few particularly striking examples.

The rare element technetium, found only as a product of uranium fission in atomic reactors, has seven valence electrons and a crystal structure favorable for superconductivity. John G. Daunt and James W. Cobble at Ohio State University had found that technetium does, in fact, become a superconductor at a relatively high temperature—11 degrees Kelvin. Now the elements before and after technetium in the periodic table, are, respectively, molybdenum and ruthenium. Ruthenium, with eight valence electrons, has to be cooled to within half a degree above zero to become superconducting. Molybdenum, with six valence electrons, cannot be made superconducting at all at the lowest attainable temperature (around one tenth of one degree above absolute zero). But if we make an alloy composed of equal parts of molybdenum and ruthenium, so that the average number of valence electrons per atom is seven and the crystal structure is identical with that of technetium, the combination becomes superconducting at 10.6 degrees—almost exactly the same as technetium's transition temperature!

Molybdenum can be made a superconductor by dissolving in it a little rhodium, which has nine valence electrons and therefore raises the average number in the mixture to slightly more than molybdenum's six. The same is true of tungsten, which like molybdenum has six valence electrons and will not become superconducting alone. We have produced many different superconductors by alloying molybdenum or tungsten with other elements, such as columbium (niobium), phosphorus, antimony or boron.

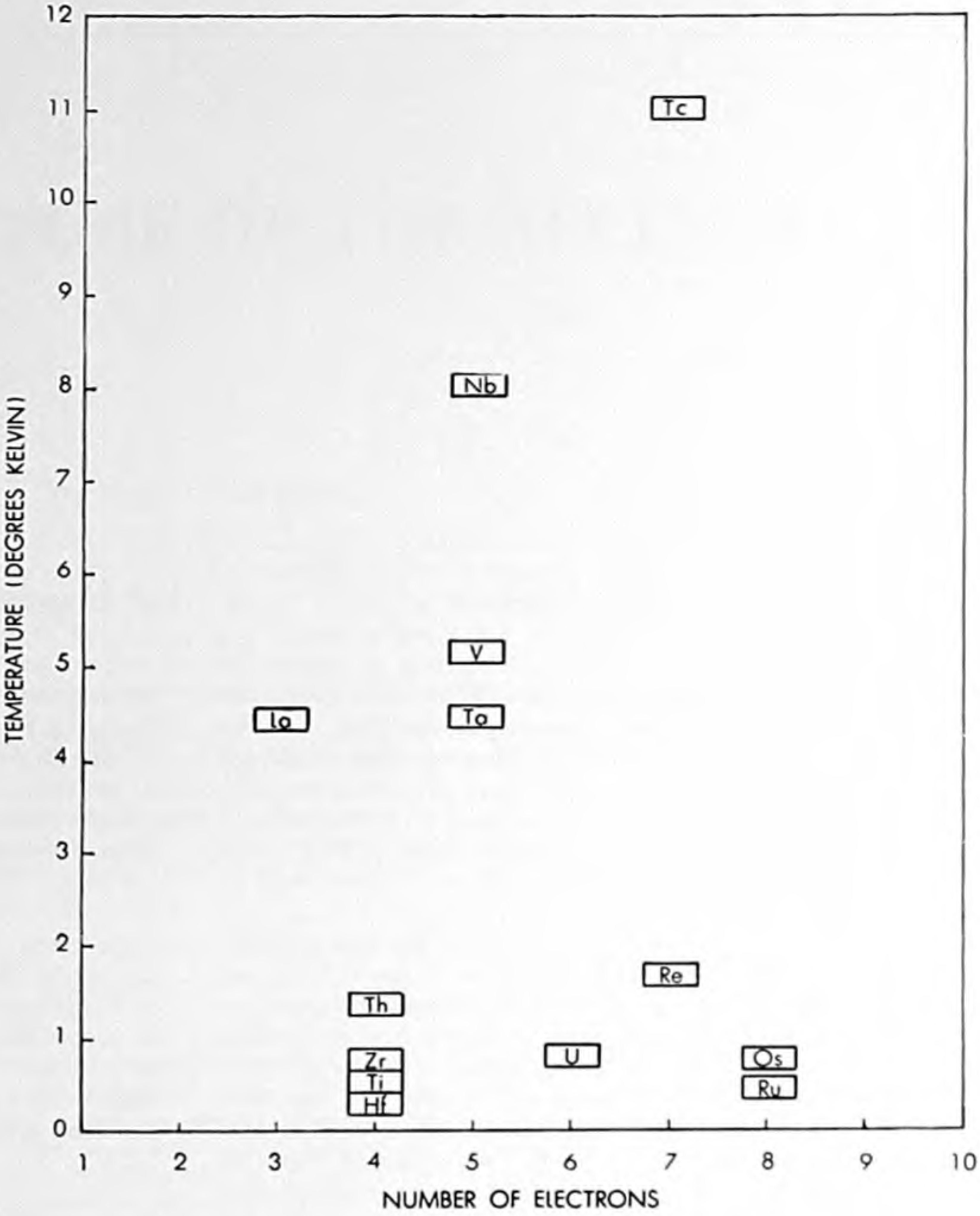
It is even possible to make a superconductor by combining two elements which by themselves are totally unfitted for this property. A good example is the combination of silicon and cobalt. Silicon, of course, is not a metal or a conductor of electricity. Cobalt has two qualities which completely disqualify it for superconductivity: it has nine valence electrons and is strongly magnetic. Yet when silicon and cobalt are combined in a cubic crystal structure, they become a superconductor, because the silicon neutralizes cobalt's magnetism and reduces the average number

of valence electrons per atom to the appropriate range.

Again, we can employ another type of manipulation based on the fact that crystal bulkiness promotes superconductivity. There is a compound of nickel and arsenic which has a favorable crystal structure and a favorable valence-electron number, yet it fails to become superconductive as we would expect. What we did in this case was to replace half the nickel with the bulkier atoms of palladium, and this mixture then became a superconductor.

We found that the most favorable crystal structure for superconductivity was the one called the beta-tungsten structure—a cubic arrangement of eight atoms which makes the crystal bulky, with considerable space between some of the atoms. Compounds with this structure and a valence-electron average between 4.5 and 4.75 per atom proved to be particularly disposed to become superconducting. On that basis we deduced that a compound of tin and columbium should reach the superconducting state at a comparatively high temperature. It turned out that this compound could be made and did in fact become a superconductor at slightly above 18 degrees Kelvin—the highest transition temperature yet found.

We can conceive of substances that may have transition temperatures as high as 30 degrees, but so far we have not succeeded in making stable compounds with the necessary specifications. At all events, transition temperatures such as 18 degrees already bring us into a range where we can begin to envision putting superconductors to work in a practicable way. Cooling apparatus for maintaining materials at these temperatures by the use of cold helium gas or liquid hydrogen is commercially available. We can now consider employing superconductors for sensitive electrical measuring instruments, switches, computer memory devices, resonating chambers in radio equipment and perhaps even transmission lines, say from the antenna to the amplifier of a radio telescope. In all these devices a superconductor would make it possible to handle extremely tiny currents or signals without any loss of their energy.



TRANSITION TEMPERATURE, or temperature at which a substance becomes superconducting, is related to the number of its valence electrons. The transition temperature of lanthanum (La) is 4.37 degrees Kelvin; of thorium (Th), 1.39 degrees; of zirconium (Zr), .7 degrees; of titanium (Ti), .53 degrees; of hafnium (Hf), .35 degrees; of niobium (Nb), 8 degrees; of vanadium (V), 5.1 degrees; of tantalum (Ta), 4.4 degrees; of uranium (U), .8 degrees; of technetium (Tc), 11 degrees; rhenium (Re), 1.7 degrees; osmium (Os), .71 degrees; ruthenium (Ru), .47 degrees. Only one group of 13 superconducting elements is shown. There is a second group of 10 elements, and a large number of compounds and alloys.

The Author

B. T. MATTHIAS is a member of the technical staff of Bell Telephone Laboratories at Murray Hill, N.J. He was born in Germany, studied physics at the University of Rome and spent the years of World War II in Switzerland, where in 1943 he earned a Ph.D. from the Federal Institute of Technology in Zurich. Until 1947 he worked as a scientific collaborator of the Federal Institute. He then came to the U.S. to join the staff of the Division of Industrial Cooperation of the Massachusetts Institute of Technology. Now a naturalized citizen of the U.S., Matthias has been with the Bell Laboratories since 1948, except for a two-year period during which he taught physics at the University of Chicago.

Bibliography

A SEARCH FOR NEW SUPERCONDUCTING COMPOUNDS. B. T. Matthias and J. K. Hulm in *The Physical Review*, Vol. 87, No. 5, pages 799-806; September 1, 1952.

SUPERCONDUCTIVITY. D. Shoenberg. Second edition. Cambridge University Press, 1952.

THE SUPERCONDUCTIVITY OF SOME TRANSITION METAL COMPOUNDS. George F. Hardy and John K. Hulm in *The Physical Review*, Vol. 93, No. 5, pages 1,004-1,016; March 1, 1954.

SUPERFLUIDS. VOL. I: MACROSCOPIC THEORY OF SUPERCONDUCTIVITY. Fritz London. John Wiley & Sons, Inc., 1950.

SCIENTIFIC AMERICAN March 1951, Vol. 184, No. 3, pp. 22-26 OFFPRINT 228

THE STRUCTURE OF THE NUCLEUS

by Maria G. Mayer

The electrons in an atom tend to occupy distinct shells. What about particles in the nucleus? The "magic numbers" of the isotope chart suggest that they also have a shell arrangement.

NO one has ever seen, nor probably ever will see, an atom, but that does not deter the physicist from trying to draw a plan of it, with the aid of such clues to its structure as he has. He needs to construct at least a rough hypothetical model of the atom as a starting point for attempting to understand its behavior. For the atom as a whole modern physicists have developed a useful model based on our planetary system: it consists of a central nucleus, corresponding to the sun, and satellite electrons revolving around it, like planets, in certain orbits. This model, although it leaves many questions still unanswered, has been helpful in accounting for much of the observed behavior of the electrons. The nucleus itself, however, is very poorly understood. Even the question of how the particles of the nucleus are held together has not received a satisfactory answer.

Recently several physicists, including the author, have independently suggested a very simple model for the nucleus. It pictures the nucleus as having a shell structure like that of the atom as a whole, with the nuclear protons and neutrons grouped in certain orbits, or shells, like those in which the satellite electrons are bound to the atom. This model is capable of explaining a surprisingly large number of the known facts about the composition of nuclei and the behavior of their particles.

Let us consider first the shell structure of the electrons, from which our nuclear model derives. The modern exploration of the atom began with the experiments of Ernest Rutherford at Cambridge University early in this century. Rutherford's main experiment consisted in shooting high-speed alpha particles, the nuclei of helium atoms, through metal foils. The physicist-writer George Gamow has likened this experiment to the strategy of an overworked South American customs inspector who adopted the expedient of shooting revolver bullets into bales of cotton as a quick method of finding out whether they contained contraband arms. (His bullets must have been shot from a revolver of very high muzzle velocity to penetrate the bales, but this small detail hardly destroys the analogy.) If the bale contained only cotton, the bullets would pass through in approximately a straight line. If, nowever, the bale contained metal arms, some bullets would ricochet off them and be deflected at a wide angle. Rutherford found in his experiment that most of the alpha particles passed through the metal foil with no appreciable deflection in angle, though of course they were slowed down. But a certain small proportion of his projectiles were markedly deflected. From this it could be concluded that the greater part of the metal foil consisted of light cottonlike material, and that the foil also contained some heavy small targets capable of exerting very strong forces on the alpha particles.

It was essentially from this experiment that the modern concept of the planetary atom emerged. The heavy targets that deflected some of Rutherford's projectiles were the nuclei of atoms. Almost all the mass of an atom is concentrated in this small, positively-charged nucleus. The diameter of the nucleus is only about a hundred thousandth that of the whole atom. The remaining space in the atom is almost empty, and in this space the electrons move in their orbits. Most of Rutherford's alpha particles never came close to a nucleus, but they encountered many of the light electrons as they passed through the outer spaces of the atoms and were thereby slowed down.

WITH the leads furnished by Rutherford's experiments and by the quantum theory, Niels Bohr and other physicists went on to establish the fundamental laws governing the motion of the electrons around the nucleus. In these studies they worked with the simple hydrogen atom, which has only a single positive charge on its nucleus and a single electron. The hydrogen atom's electron may move in any one of a discrete series of elliptical orbits about the nucleus. Under the quantum rule only these specific orbits are stable, or "permitted," for the electron. The most stable of the allowed orbits is that of lowest energy; in this orbit the electron keeps within about one Angstrom unit (1/100,000,000 of a centimeter) of the nucleus. Here its "angular momentum" is zero. Angular momentum is a measure of the amount of rotation of a particle in an orbit; for a circular orbit it is computed as the mass times the velocity times the radius of the orbit. When measured in appropriate units ($h/2\pi$, with h representing Planck's constant), the angular momentum of a permitted Bohr orbit is always a whole number. At the next higher level of energy an electron is allowed any one of four possible orbits. One of these has zero angular momentum; the other three all have an angular momentum of one unit but differ in the direction of the angular momentum vector. At still higher energy levels the electron may have more than four allowed orbits.

Besides revolving around the nucleus, the electron also spins on its own axis, just as the earth has a daily rotation in addition to its motion around the sun. The electron's spin can take one of two directions. This effectively doubles the number of permitted motions, or kinds of orbit, for an electron, since in any

given orbit it may spin in one direction or the other. The angular momentum of the electron's own spin is a half-unit.

In an atom heavier than hydrogen, *i.e.*, with more than one electron, the electrons' motions are of course more complicated. Just as the orbit of the earth around the sun is distorted from a perfect ellipse by the gravitational effects of the other planets, so the motion of each electron around the nucleus of a heavy atom is influenced by the presence of the other electrons. To some extent, however, these effects can be averaged and the motion of a single electron can be considered to be governed by the averaged field of force from the nucleus and the other electrons. A further factor that must be noted is the famous Pauli exclusion principle: namely, that two electrons can never exist in exactly the same orbit of the same magnitude and direction of orbital angular momentum and with their spins in the same direction.

WE might liken the electronic structure of an atom to that of an apartment house. The ground floor, or orbit of lowest energy, which is the most desirable from the electron's point of view, has a single apartment. Two electrons of opposite spin can occupy it. The second floor contains four apartments, one of which is at a slightly lower level than the others. This one has zero angular momentum, while the other three have one unit of angular momentum. Each of these four apartments likewise can be occupied by two electrons of opposite spin. Thus the apartment house rises, level by level, with a certain number of apartments at each level, each available for occupancy by a pair of electrons. The atomic physicist has designated the levels (or orbits or shells or whatever one pleases to call them) by a system of numbers and letters: the letters represent the energy levels in terms of angular momentum. The letter s stands for zero angular momentum; p for one unit of angular momentum, d for 2, f for 3, and so on. Thus the ground floor, the lowest orbit that can be occupied by the hydrogen atom's electron, with zero angular momentum, is called 1s; the next orbit, that of the depressed second-floor apartment with zero angular momentum, is 2s; the next, occupied by the three second-floor apartments with one unit of angular momentum, is 2p, and similarly up through the higher orbits.

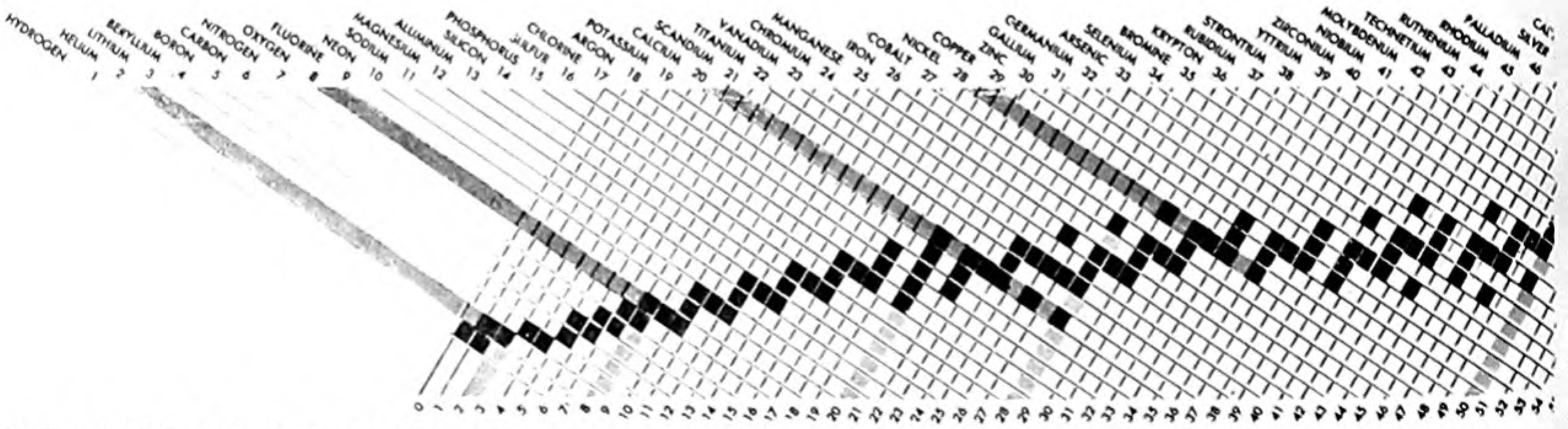
Now it turns out that atoms are happiest and most stable when all the apartments at a given level are fully occupied by electrons; they do not like vacancies. For example, the single electron of the hydrogen atom is lonely and restless in its 1s apartment, which as we have seen has room for two electrons. Hydrogen readily loses its electron, thereby becoming a positive ion, or picks up a roommate for its lone electron, forming a negative ion. Hydrogen, in other words, is a very reactive element; it easily enters into chemical combinations with other atoms. In contrast, the helium atom, with two electrons, is extremely stable. Its two electrons fill the 1s apartment and form a closed system. There is no room for more electrons in this orbit, and the two electrons are held so tightly by the two positive charges in the atom's nucleus that they are not easily separated from it. As a result helium does not react or combine with other elements. The same is true of the four other "noble" gases in the periodic table of elements—neon, argon, krypton and xenon. Each of these has a closed outer shell, fully occupied by electrons, and does not form ions easily or enter into chemical combinations. Indeed, the stability of the five noble gases is so remarkable that their atomic numbers—2, 10, 18, 36 and 54—may be called the "magic numbers" of the periodic table.

This picture of the atom's electronic structure has been enormously useful to chemistry; by means of it chemists have been able to group the elements in families and predict their chemical behavior. But it tells us little or nothing about the structure of the nucleus. The only piece of information an atom's chemical activity gives us about its nucleus is the number of positive charges, or protons, it contains. The chemical nature of the atom is determined by the number of its electrons. We know that if an atom has a certain number of electrons, it must have the same number of protons in its nucleus to hold them. But how the protons themselves are held together, how they interact with one another and with the uncharged neutrons also present in the nucleus—these are the great mysteries of nuclear physics.

The forces that bind the nucleus together are very great—millions of times greater than those that bind electrons to the nucleus. We do not know how these forces are created. And even if we understood them completely, we would still be confronted with the tremendous difficulty of solving a many-body problem, *i.e.*, calculating the results of these forces upon a large number of protons and neutrons that interact with one another strongly and at extremely short range within the nucleus. The problem has seemed so hopeless that nuclear physicists have preferred to treat the nucleus as a "liquid drop," in which the protons and neutrons essentially lose their identity. (This approach has been very fruitful, particularly in studying high-energy phenomena in the nucleus.)

Yet it is possible to discern some rather remarkable patterns in the properties of particular combinations of protons and neutrons, and it is these patterns that suggest our shell model for the nucleus. One of these remarkable coincidences is the fact that the nuclear particles, like electrons, favor certain "magic numbers."

EVERY nucleus (except hydrogen, which consists of but one proton) is characterized by two numbers: the number of protons and the number of neutrons. The sum of the two is the atomic weight of the nucleus. The number of protons determines the nature of the atom; thus a nucleus with two protons is always helium, one with three protons is lithium, and so on. A given number of protons may, however, be combined



THE CHART OF THE ISOTOPES reveals seven "magic numbers" (*rows shaded red and black*). Each isotope is represented by a black square. Each row outlined by black lines includes nuclei with the same number of protons, *i.e.*, isotopes; each row enclosed in red lines shows nuclei with same number of neutrons. Since

with varying numbers of neutrons, forming several isotopes of the same element. Some isotopes are stable; others decay by radioactivity. Some of the stable isotopes readily add a neutron; others are much less inclined to do so.

Now it is a very interesting fact that protons and neutrons favor even-numbered combinations; in other words, both protons and neutrons, like electrons, show a strong tendency to pair. In the entire list of some 1,000 isotopes of the known elements, there are no more than six stable nuclei made up of an odd number of protons and an odd number of neutrons. The other odd-odd nuclei break down radioactively by emitting a negative or positive electron; this change in charge transforms a neutron into a proton or a proton into a neutron and creates a more stable even-even combination of protons and neutrons.

Moreover, certain even-numbered aggregations of protons or neutrons are particularly stable. One of these magic numbers is 2. The helium nucleus, with 2 protons and 2 neutrons, is one of the most stable nuclei known. The next magic number is 8, representing oxygen, whose common isotope has 8 protons and 8 neutrons and is remarkably stable. The next magic number is 20, that of calcium. Calcium, with 20 protons, has 6 stable isotopes, ranging in neutron number from 20 to 28. This is an unusually large number of stable isotopes for the lower region of the periodic table.

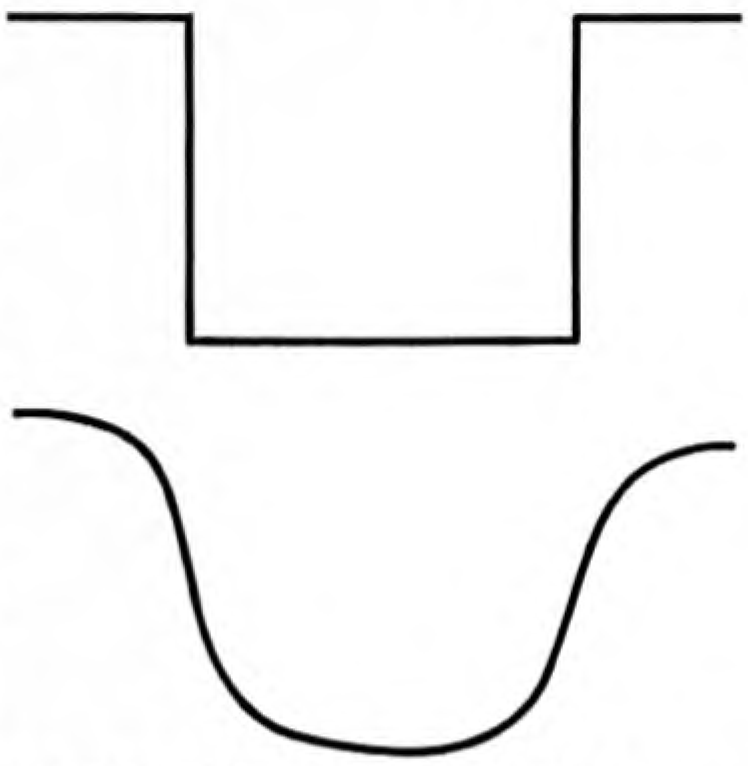
Among these light elements the relative stability can be determined very accurately in terms of binding energy. The net mass of a nucleus is always smaller than the combined masses of the protons and neutrons of which it is composed. The binding energy is calculated from this "mass defect" by means of Einstein's famous relation $E=mc^2$, with m representing the mass defect and c the velocity of light. Such calculations show conclusively that the nuclei with the magic numbers 2, 8 and 20 have much greater binding energies than their neighbors. But for the heavier elements above calcium the binding energies are not accurately determined, and we must judge their relative stability by indirect evidence. One such piece of evidence is the number of stable (*i.e.*, nonradioactive) nuclei that are found to exist with a given number of protons or neutrons. Another is the relative abundance of a given nucleus in the universe, since it seems reasonable to assume that the most abundant isotopes are the most stable.

By these tests the number 50 joins the list of magic numbers. Tin, with 50 protons, has 10 stable isotopes, more than any other element, and it is much more abundant than neighboring elements in the periodic table. The same is true, to a somewhat lesser degree, of the number 28. Another magic number is 126: an isotope with 126 neutrons holds them much more strongly than one with 127 or 128. Perhaps the most remarkable magic number of all is 82. There are 7 stable nuclei containing 82 neutrons, ranging from isotopes of xenon to samarium. The barium isotope with 82 neutrons accounts for 72 per cent of the abundance of that element, and cerium's 82-neutron isotope represents 88 per cent of all the cerium. Finally, 82 protons means lead, and lead is the stable end-product of the decay of all the heavy radioactive elements that may be found in nature.

There are other indications of the special stability of these magic numbers. For instance, nuclei containing 50, 82 or 126 neutrons do not like to add an extra neutron: their absorption cross-sections for fast neutrons are smaller by several factors of 10 than those of an average nucleus of nearly the same weight.

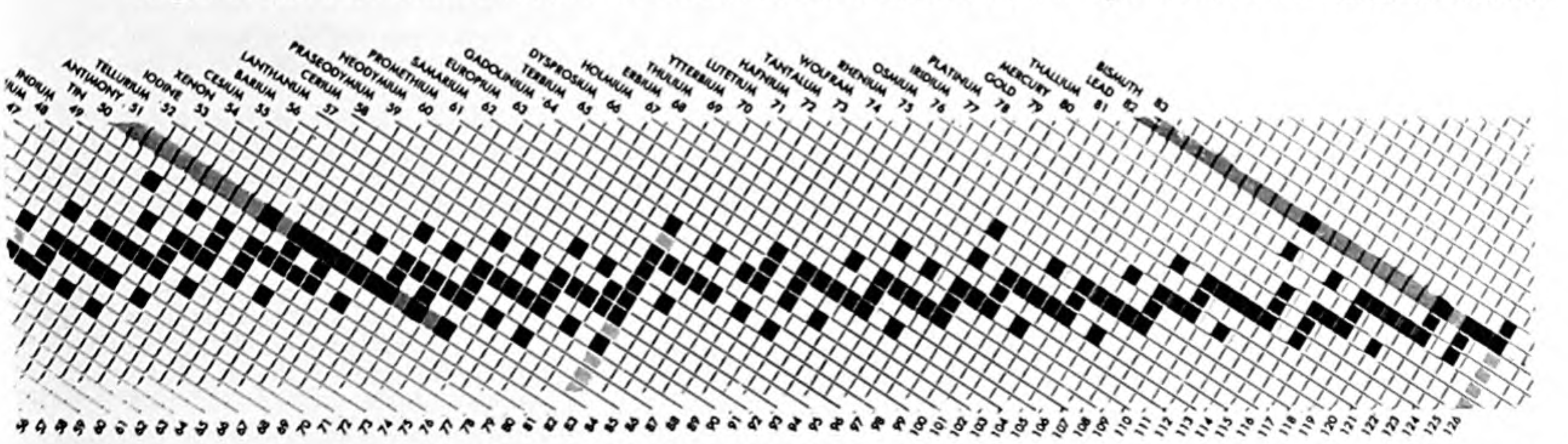
THE list of magic numbers, then, is: 2, 8, 20, 28, 50, 82 and 126. Nuclei with these numbers of protons *or* neutrons have unusual stability. It is tempting to assume that these magic numbers represent closed shells in the nucleus, like the electronic shells in the outer part of the atom. To be sure, the electronic shells form a more distinct dividing line than those in the nucleus: the last electron in a closed shell is held to the atom at least three times as strongly as the last electron in an unfilled shell, whereas in nuclei the energy difference is at most 50 per cent. Yet the situations seem sufficiently similar to justify exploring the possibility that the nucleus may be tied together in much the same kind of shell structure as the electrons.

There is another kind of evidence that



NUCLEAR POTENTIALS are represented as "wells." A well with round edges agrees with magic numbers better than one with square.

supports this shell-structure hypothesis. It has to do with the spin of the nucleus. Many nuclei apparently spin about their axes like a top, just as the earth and electrons do. Since the nucleus carries a charge, its rotation corresponds to an electric current, and it behaves like a tiny magnet. The result is known as the magnetic moment of the spinning body. The spins and the magnetic moments of nuclei and of their building blocks, the neutron and proton, can be measured. It turns out that the spin of nuclei, like that of electrons, is "quantized": that is, when measured in the same units ($h/2\pi$) it is always a whole number or half a whole number. The spin of the proton and of the neutron has been de-



the number of protons is the same as the atomic number, the name of each element is given above the number of protons. The nuclei that contain 2, 8, 20, 28, 50, 82 or 126 protons *or* neutrons are unusually stable: these are the magic numbers. The chart is not carried beyond bismuth: that is the last element to which magic numbers apply.

ORBITAL ANGULAR MOMENTUM	SHELL	NUMBER OF NUCLEONS IN A LEVEL	NUMBER OF NUCLEONS UP TO A LEVEL
0	1s	2	
			2
1	1p	6	
			8
2	1d	10	
0	2s	2	
			20
3	1f	14	
1	2p	6	
			40
4	1g	18	
2	2d	10	
0	3s	2	
			70
5	1h	22	
3	2f	14	
1	3p	6	
			112
6	1i	26	
4	2g	18	
2	3d	10	
0	4s	2	
			168

MARGENAU CHART of nuclear energy levels supports the shell-structure hypothesis in part. The first three numbers in column at right match the magic numbers. The next four do not; at this point the scheme breaks down.

ORBITAL ANGULAR MOMENTUM	SHELL	TOTAL ANGULAR MOMENTUM	NUMBER OF NUCLEONS IN A LEVEL	NUMBER OF NUCLEONS UP TO A LEVEL
0	1s	1/2	2	
				2
		3/2	4	
1	1p	1/2	2	
				8
		5/2	6	
2	1d	3/2	4	
0	2s	1/2	2	
				20
		7/2	8	
				28
3	1f	5/2	6	
		3/2	4	
1	2p	1/2	2	
		9/2	10	
				50
4	1g	7/2	8	
		5/2	6	
2	2d	3/2	4	
0	3s	1/2	2	
		11/2	12	
				82
5	1h	9/2	10	
		7/2	8	
3	2f	5/2	6	
		3/2	4	
1	3p	1/2	2	
		13/2	14	
				126
6	1i	11/2	12	

MAYER CHART accounts for the discrepancies by introducing the spins of nuclei (*fractions in the center column*). "Spin-orbit coupling" splits the close-lying levels apart and creates energy gaps where the magic numbers occur.

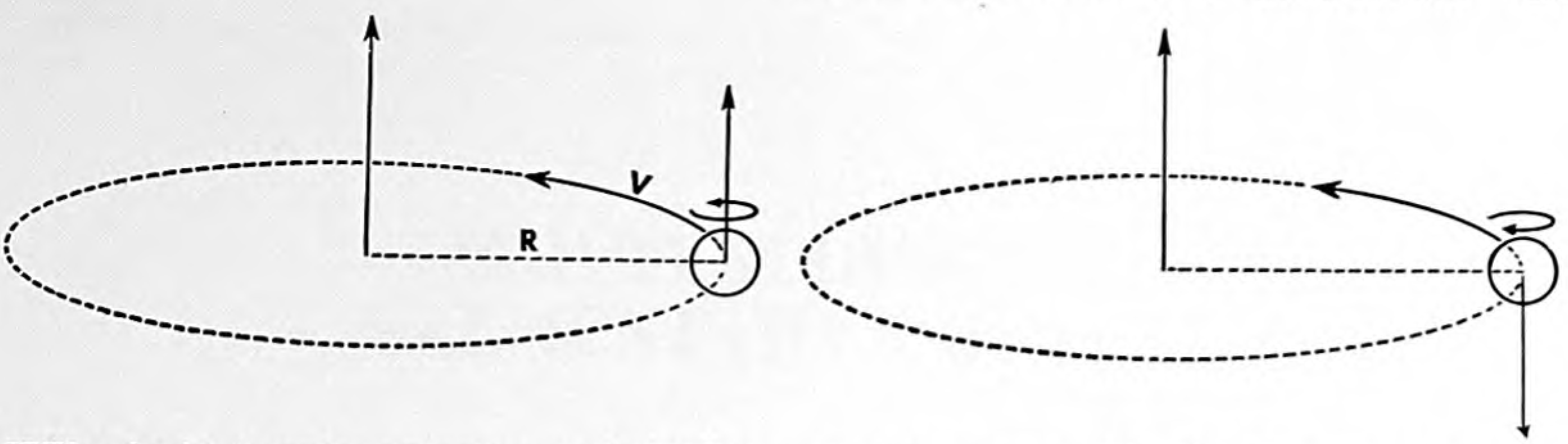
termined to be 1/2-unit. Their magnetic moments, measured in units called nuclear magnetons, also have been accurately determined. The spins of nuclei with an odd number of particles are all half of some whole odd number (*e.g.*, 1/2, 3/2, 5/2 and so on), and the spins and magnetic moments of nuclei with an even number of protons and of neutrons are zero.

Now it is a very surprising fact, not expected from the general theory of the nucleus, that two isotopes with the same odd number of protons but different even numbers of neutrons behave very similarly. They have the same spins, nearly the same magnetic moments and frequently the same kinds of excited states. Take, for example, the 2 isotopes indium 113 and indium 115. Each has 49 protons, but one has 2 neutrons more than the other. Yet both have spins of 9/2, and their magnetic moments are very close in value—5.461 and 5.475 nuclear magnetons, respectively. The extra pair of neutrons in the second isotope does not seem to affect these nuclear properties; in other words, the spins and magnetic moments in this case appear to be due only to the protons. On the other hand, a nucleus with 49 neutrons and an even number of protons (*e.g.*, strontium-87) also has a spin of 9/2, though not the same magnetic moment. As far as nuclear spin goes, 49 neutrons behave just like 49 protons. This seems to be the general rule for the lighter isotopes of mass number less than 120.

On the basis of these observations the German physicist T. Schmidt many years ago made the radical suggestion that it is not the nucleus as a whole that spins. Instead, in nuclei with an odd number of particles the properties of spin and magnetic moment are due entirely to the last odd particle, be it a proton or a neutron. The structure of odd nuclei is thus pictured very simply. The nucleus has a spherically symmetrical core of an even number of neutrons and protons. The core has no spin. Around it revolves the last odd particle. Its motion alone determines the spin and magnetic moment of the nucleus.

It is curious to find that, as the evidence of the magic numbers and of the nuclear spins suggests, protons and neutrons behave with considerable independence of each other in the nucleus. Such independence is unexpected from the standpoint of current nuclear theory.

The angular momentum of a particle's orbit alone is always a whole number. But the particle's total angular momentum is the sum of its orbital momentum and its own spin. Hence the total angular momentum of a nucleus always has a half-integer value, since protons and neutrons, just as electrons, spin about their own axes with an angular momentum of 1/2. This spin can be parallel or anti-parallel to the orbital angular mo-



SPIN and orbital angular momentum are depicted in this diagram. The spin is indicated by the circular arrow above the particle. The orbital angular momentum is the mass of the particle times R times V.

mentum; that is, it can be directed so as to add 1/2 or to subtract 1/2 from the orbital angular momentum. Consequently a measured spin can correspond to either of two different orbits.

That such a picture is not too far from the truth is borne out by the measured values of magnetic moments. From this simple model it is possible to compute the theoretical magnetic moment of a given nucleus. The agreement between the predicted and measured magnetic moments is reasonably good.

THUS the spins and magnetic moments lead to a description of the nucleus in terms of orbits of single particles. One can then picture the building of the structure of the nucleus as the gradual filling up of single-particle orbits by neutrons and protons, in the same way as electrons build the atom. The single-particle orbits for one nuclear particle are determined by the average field of the others. The orbits can be described by the letters used to designate the quantized orbital angular momentum of electrons: s=0, p=1, d=2, f=3, g=4, h=5, i=6. Since neutrons and protons obey the Pauli principle, the s level has room for just two protons and two neutrons, as in the case of electron orbits; the p level has room for six, and so on. In this scheme the magic numbers should correspond to closed shells; that is, they should indicate the boundaries where one level is filled and the next level is appreciably higher.

The order of levels, or the number of particles in a level, depends on the form of the nuclear potential. The simplest guess we can make about the potential is to compare it to a well: no force is acting on a particle outside the well (*i.e.,* outside the spherical nucleus), and none inside. But at the edge of the well (at the surface of a spherical shell) a strong attraction takes place. The change in attraction may be more or less abrupt. A series of very abrupt transitions would be represented by a well with square edges; a more gradual transition by a well with rounded edges.

The structure of levels that would obtain if the well had square edges has been calculated by the Yale University physicist Henry Margenau. The result is shown in the table at the top of page 230. It gives the number of neutrons or protons each level can hold and the cumulative numbers that fill all the levels up to a given point. The same type of calculation for wells with rounded edges divides each level into two or more levels of the same or nearly equal energy; in the table these closely adjacent levels are grouped together, with breaks between the gaps.

This scheme explains perfectly the smaller magic numbers, up to 20. The lowest level can contain no more than two particles. The next level, or p-shell with angular momentum 1, has three different orbits, each of which can be occupied by two nuclear particles with opposite spin. These six particles in the p-shell, plus the two in the ground state, correspond to the stable form of oxygen with eight neutrons and eight protons. The next two levels, 1d and 2s, which lie close together, have room for 12 more nuclear particles, and these bring us to the magic number 20.

From here on the scheme seems to break down. In this table there is no sign of the magic numbers 28, 50, 82 and 126. It is possible, however, to arrive at them by taking into account an effect which is very small and unimportant for electrons but apparently not for nuclear particles.

A nuclear particle can align its own 1/2-unit spin parallel or anti-parallel with the angular momentum of its orbit. Thus its total angular momentum can be the orbital angular momentum plus 1/2 or minus 1/2. This splits its orbit into two possible levels. There is a considerable difference of energy between these two levels, brought about by what is called "spin-orbit coupling." The strength of the coupling increases with increasing angular momentum, and one would expect it to decrease with increasing size of the nucleus. Consequently the greatest split in such a pair of levels should occur at the beginning of a group of closely adjacent shells, where relatively high angular momentum is combined with a relatively small nucleus. Suppose that a wide split of this kind occurs at the 1g level. Then it is no longer correct to treat the 1g level as a single level with room for 18 particles. Instead, it will divide into two levels with room for 10 particles in its lower level. This lower level will be depressed toward the next lower group, consisting of the 1f and 2p shells, as listed in the Margenau table. The 10 particles, added to the 40 up to that point, make the magic number 50. Above this there is an energy gap, created by the spin-orbit coupling.

The table at the bottom of page 230 shows how the various levels are revised when allowance is made for spin-orbit coupling. Now the larger magic numbers come into the scheme: they all occur at the places where spin-orbit coupling has its greatest effect.

This level scheme is in excellent agreement with the observed spins and magnetic moments of odd-numbered nuclei. Among the approximately 90 spins and 73 magnetic moments that have been measured so far there are only four serious disagreements with the theory.

THE shell model can explain other features of nuclear behavior, including the phenomenon known as isomerism, which is the existence of long-lived excited states in nuclei. Perhaps the most important application of the model is in the study of beta-decay, *i.e.*, emission of an electron by a nucleus. The lifetime of a nucleus that is capable of emitting an electron depends on the change of spin it must undergo to release the electron. Present theories of beta-decay are not in a very satisfactory state, and it is not easy to check on these theories because only in a few cases are the states of radioactive nuclei known. The shell model can help in this situation, for it is capable of predicting spins in cases in which they have not been measured.

Certainly the simple model described here falls short of giving a complete and exact description of the structure of the nucleus. Nonetheless, the success of the model in describing so many features of nuclei indicates that it is not a bad approximation of the truth.

The Author

MARIA G. MAYER is a physicist with the Argonne National Laboratory (AEC) and the Institute of Nuclear Studies of the University of Chicago. She was born in Kattowitz, Poland, and was educated at the University of Göttingen, in Germany. There she met her husband, Joseph E. Mayer, an International Education Board Research Fellow who was spending a year of post-graduate study at the University. Upon receiving her doctorate she emigrated to the U. S. with her husband in 1930. They both held appointments at Johns Hopkins, were parted academically when he taught for a few years at Columbia and she at Sarah Lawrence College, and then were reunited at the University of Chicago. The Mayers have two children, are members of the National Academy of Sciences and have collaborated on a book on statistical mechanics. Maria Mayer is distinguished for her contributions to quantum mechanics and to the subject of this article.

Bibliography

NUCLEAR CONFIGURATIONS IN THE SPIN ORBIT COUPLING MODEL. Maria G. Mayer in *The Physical Review*, Vol. 78, No. 1, pages 16-23; April 1, 1950.

NUCLEAR SHELL STRUCTURE. Eugene Feenberg and Kenyon C. Hammack in *The Physical Review*, Vol. 75, No 12, pages 1877-1893; June 15, 1949.

ON SPINS, MOMENTS, AND SHELLS IN NUCLEI. L. W. Nordheim in *The Physical Review*, Vol. 75, No. 12, pages 1894-1901; June 15, 1949.

SCIENTIFIC AMERICAN April 1958, Vol. 198, No. 4, pp. 56-64 OFFPRINT 229

THE TEACHING OF ELEMENTARY PHYSICS

by Walter C. Michels

To give more students a better understanding of the nature of physical science, a new approach to high school physics is evolving. Its main feature: emphasis on basic principles.

In the public discussion that has followed the Russians' launching of their satellites, the word "physics" has probably occurred almost as often as "satellite," "rocket" or "missile." Many speakers and writers have implied that the technological progress of the U.S.S.R. can be attributed entirely to the fact that even high-school graduates there have had some 12 years of study in mathematics and six in physics. It seems to follow that all the U. S. needs to do to regain undisputed scientific and technological leadership in the world is to introduce more physics courses and mathematical instruction into its schools. Like all panaceas, this cure for our ills has the attraction of simplicity. It has the further attraction of potent medicine, because physics and mathematics are considered "hard" subjects and we still retain some of the atavistic belief that unpleasant medicines are especially effective—particularly when our children, rather than we, take them.

But before we rush into a program designed only to persuade more students to take more physics, it may pay us to examine physics instruction in the context of American education as a whole, to ask what we want it to accomplish and to estimate whether improvement in quantity or in quality is the more important.

At present, between one fifth and one fourth of the students graduating from high schools in the U. S. have taken a course in physics. This is almost the same as the proportion of high-school graduates who go on to college. The agreement is more than coincidental—most of the leading colleges either require or strongly encourage some preparation in physics. The engineering schools, which are the first choices of about one quarter of all the boys in the college preparatory group, almost invariably require physics for admission. Thus a large percentage of the students who take physics in high school are a "captive audience," in the sense that they are required to pass the course to get into college.

Much of the argument for bringing more students into secondary-school physics courses seems to be based on the idea that this will ultimately produce more scientists and engineers. It is difficult to predict the extent to which this recruiting effort, even coupled with offers of college scholarships, will succeed. Motivational factors and temporary fluctuations in the demand for professional workers may well be more powerful influences than educational offerings in shaping the careers of young men and women. Some increase in the number of available engineers and scientists will undoubtedly result from present pressures and from Federal support of education; yet we may be disappointed in the results if we set this increase as our only, or even our primary, goal.

Many scientists and some of our best educators believe that there is a more important reason to increase emphasis on physics in the schools. They believe that our educational system is failing to give most of our population an adequate understanding of the nature of physical science and of its role in the economic, political and cultural life of the 20th century. George Stewart, in his book *Man—An Autobiography*, points out that mankind's way of life has changed more since 1700 than it had from prehistoric times to that date. Our economy of abundance, our increased leisure with its cultural benefits, our rapid communication systems with their effect on international relations, our increasing ability to employ nature for our own ends—all of these are direct consequences of the ideas and the methods introduced by Galileo, Newton and the physical scientists who have stood on the shoulders of these giants. An understanding of science is as important to the businessman, to the lawyer, to the statesman or to any other citizen as an understanding of language or literature or of political and military history. Physics is the most basic of the natural sciences. For this reason, the study of physics is essential to the education of all of the population. The future rests not so much on the number of engineers and scientists produced each year as it does on the intellectual and cultural climate in which they work, and this, in turn, will be determined by our success or failure in making physical science a significant part of general education.

Today's Physics Courses

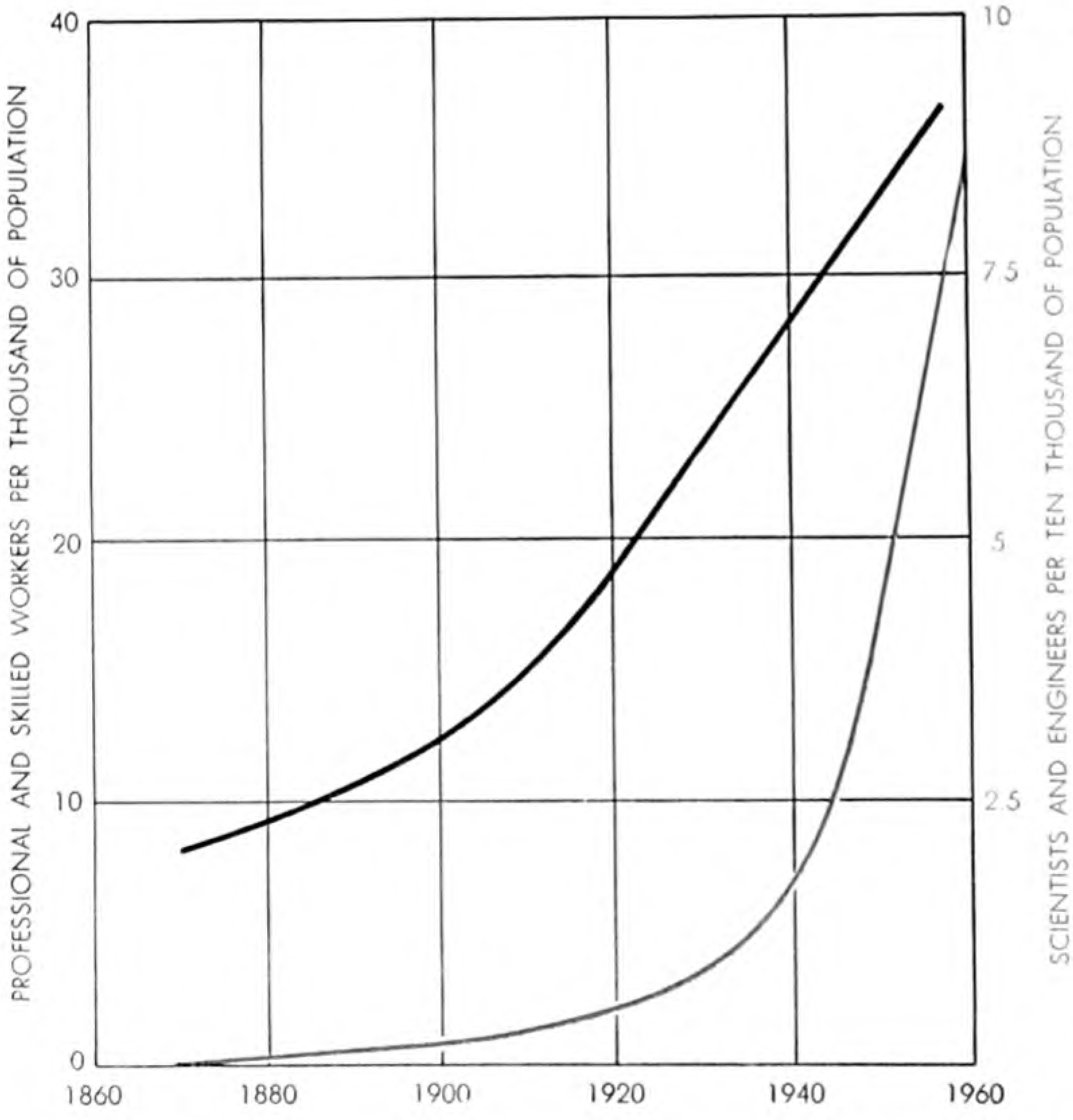
Physics, like any other subject, can contribute whatever the student is willing and able to take—anything from true understanding down to mastery of a few facts to be regurgitated on an examination. The quality of the course itself will depend on the textbook, on the laboratory aids available and, most of all, on the ability of the teacher. A truly able

teacher possesses both a knowledge that goes well beyond the level at which he is teaching and an understanding of the people he is teaching. From the first of these he draws illustrative materials and constantly changing methods of presentation, and so makes his subject interesting and exciting enough to fire the imagination of his students. The second quality allows him to tailor his approach to the mental abilities of the students and to choose the phrase or the illustration that will make his meaning clear or will untangle a difficult passage in the textbook. It is because both of these abilities are needed that we are faced with so serious a shortage of science teachers. We cannot solve the problem by bringing into the classroom scientists with an advanced knowledge of their subject but with no feeling for teaching, or by assigning physics to good teachers who know little physics.

In the best of our schools—those that attract good teachers and give them the equipment and time required for a good job—the physics courses contribute to the student's education many things that he cannot obtain from other studies. Here he learns the basic principles that underlie events from the throwing of a baseball to the launching of a satellite; he learns to observe phenomena with care, to extract from them the parts that are most important and to use his observations to test a theory. Because the phenomena dealt with are simple and subject to measurement, he obtains practice in the analysis of their causes—practice which will stand him in good stead when he later faces the problem of analyzing more complicated happenings without allowing his prejudices to influence his interpretation of them. His interest in and understanding of algebra and geometry are likely to be advanced as he sees that these subjects supply valuable tools for the description both of nature and of the products of modern technology. He learns to read carefully, with the idea that every word in a well-written piece of exposition is there for a purpose; at the same time, he learns to communicate ideas in a clear, succinct and unambiguous manner, using graphic methods and mathematical descriptions as well as words. Whether the student is destined to become an engineer, a businessman, a lawyer or a housewife, he or she can be expected to retain these advantages of the physics course long after the detailed facts and the techniques of physics are forgotten.

Few physics courses accomplish all of these things, and many fall far short of the goals that we have set. One reason for their shortcomings is that tradition, economic pressure and mistaken ideas about the nature of science have given us textbooks which sometimes hinder rather than help the teacher. The content of the courses is heavily influenced by the fact that a large proportion of the students taking them are preparing for engineering. These students are interested in the technical applications of physics. The authors of textbooks and the teachers are tempted to catch their interest by loading the course with technology. Some of the textbooks abound in pictures and diagrams of Diesel shovels, locomotives, gasoline engines and automobile transmissions. No one can object to the use of practical machines to illustrate the principles of physics, but we must protest when the text is so overloaded with these illustrations that the principles themselves have to be treated in a summary fashion, if not omitted entirely. Further, many modern machines are too complex to be described adequately or discussed correctly on the basis of the physical knowledge available to the students. Under these circumstances, too many physics courses demand only rote memorization of isolated facts and of formulas into which numbers can be substituted to achieve answers which often mean little to the student who has obtained them.

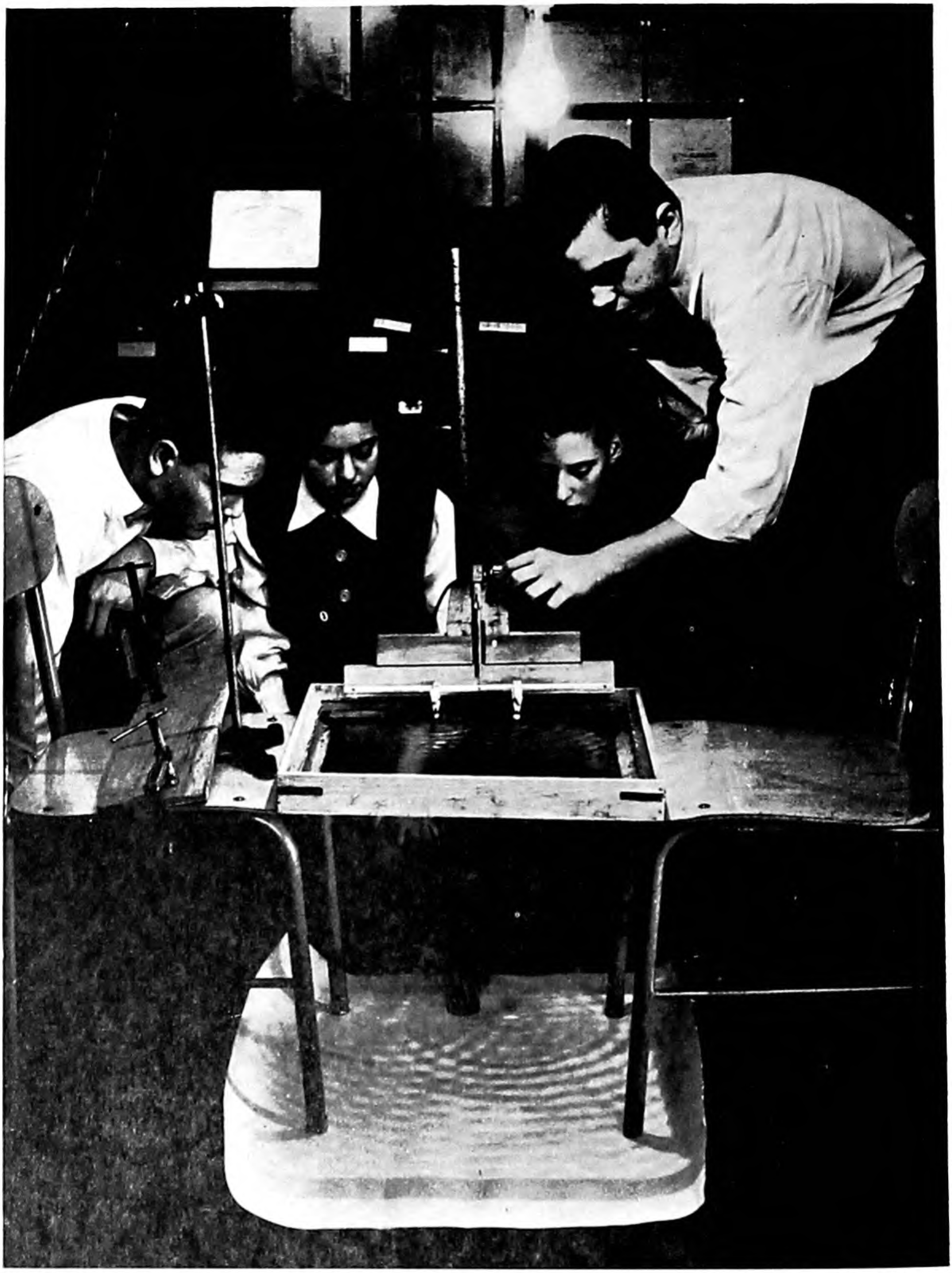
A second fault of many textbooks is the breakdown of physics into topics or "units"—a misuse of the basically sound pedagogical idea that students should



PROGRESS OF CIVILIZATION requires an increasing proportion of technically trained individuals in the population. The graph shows the historical rate of growth of this group.

UNITY OF PHYSICS is reflected in the drawing on the opposite page, which shows how the branches of physics have converged to form the subject that must be taught today. The pioneers whose portraits appear here are: Galileo Galilei (1), Isaac Newton (2), Count Rumford (3), James Prescott Joule (4), Lord Kelvin (5), Joseph Black (6), René Descartes (7), Christian Huygens (8), Thomas Young (9), Augustin Jean Fresnel (10), Benjamin Franklin (11), André Marie Ampère (12), William Gilbert (13), Hans Christian Oersted (14), Michael Faraday (15), James Clerk Maxwell (16), Hendrik A. Lorentz (17), Max Planck (18) and Albert Einstein (19).


Mechanics
Heat
Light
Electricity
Magnetism
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.
13.
14.
15.
16.
17.
18.
19.
BERNARDA BRYSON



RIPPLE TANK, in which a pair of vibrating plungers generate sets of circular waves, is used to study wave phenomena in the laboratory. The apparatus pictured here and on the following three pages was developed by the Physical Science Study Committee. Much of the material is built by the students themselves. The photographs were made at the Bronx High School of Science in New York. This is one of the schools where the Committee's new program is being tested. The instructor of the course is Alexander Joseph.

not be confused with too many facts or concepts at any one time. The difficulty is that the texts rob the units of meaning. Sound, heat, light, electricity, magnetism, mechanics and atomic physics are discussed separately as if they had little or no connection with one another. This method of presentation belies the very nature of the subject, for the chief thing that distinguishes the physical sciences from other intellectual activities is the existence of a central theory, based on a very few simple laws and principles, which applies to all phenomena. All of the particular subjects listed above can be closely related on the basis of fundamental principles such as the conservation of energy, of matter and of momentum.

Principles *v.* Applications

As an example of the failure of the piecemeal approach let us take one of the bulwarks of physics courses—Ohm's law. In the first half of the 19th century Georg Simon Ohm of Germany showed that the voltage required to produce a steady current in any part of a given electric circuit was proportional to the current. That is to say, if the current maintained in a wire is to be doubled, the battery must supply twice as strong an "electrical push." Ohm thus arrived at the concept of electrical resistance and at the idea that the resistance of a conductor depended only on the material and dimensions of the wire, not on the amount of current flowing in it. We now know that Ohm's law holds only for special classes of electrical conductors (*e.g.*, metals). In fact, it is the failure of Ohm's law to apply to the flow of electricity in vacuum tubes and in the materials known as semiconductors that has made all of modern electronics possible. Yet the typical introductory physics course leaves a student with the firm conviction that Ohm's law is one of the most basic principles of physics, or at least that he will not pass his final examination if he cannot substitute numbers in the formula: $E = IR$. If he remembers that E is a voltage, I is a current and R is a resistance, he will be able to arrive at the correct answers to conventional examination questions. But if he tries to apply Ohm's formula to any circuits other than the simple ones in which he has been drilled, he will probably go far astray, for he has learned only the letter of the law, not its meaning.

Only a few years before Ohm's law was discovered, the great principle of the conservation of energy (the first law of thermodynamics) had been established. James Prescott Joule of England went on to show that heat, such as that present in a kettle of hot water, and mechanical energy, such as that carried by a rolling stone, were one and the same thing. Further, he found that electrical energy too was connected with these forms of energy. He arrived at the conclusion that a current in a wire produced heat at a rate which was proportional to the square of the current. In the same symbols that we have used to state Ohm's law, we can write Joule's law as $H = I^2R$, where H stands for the rate at which heat is produced.

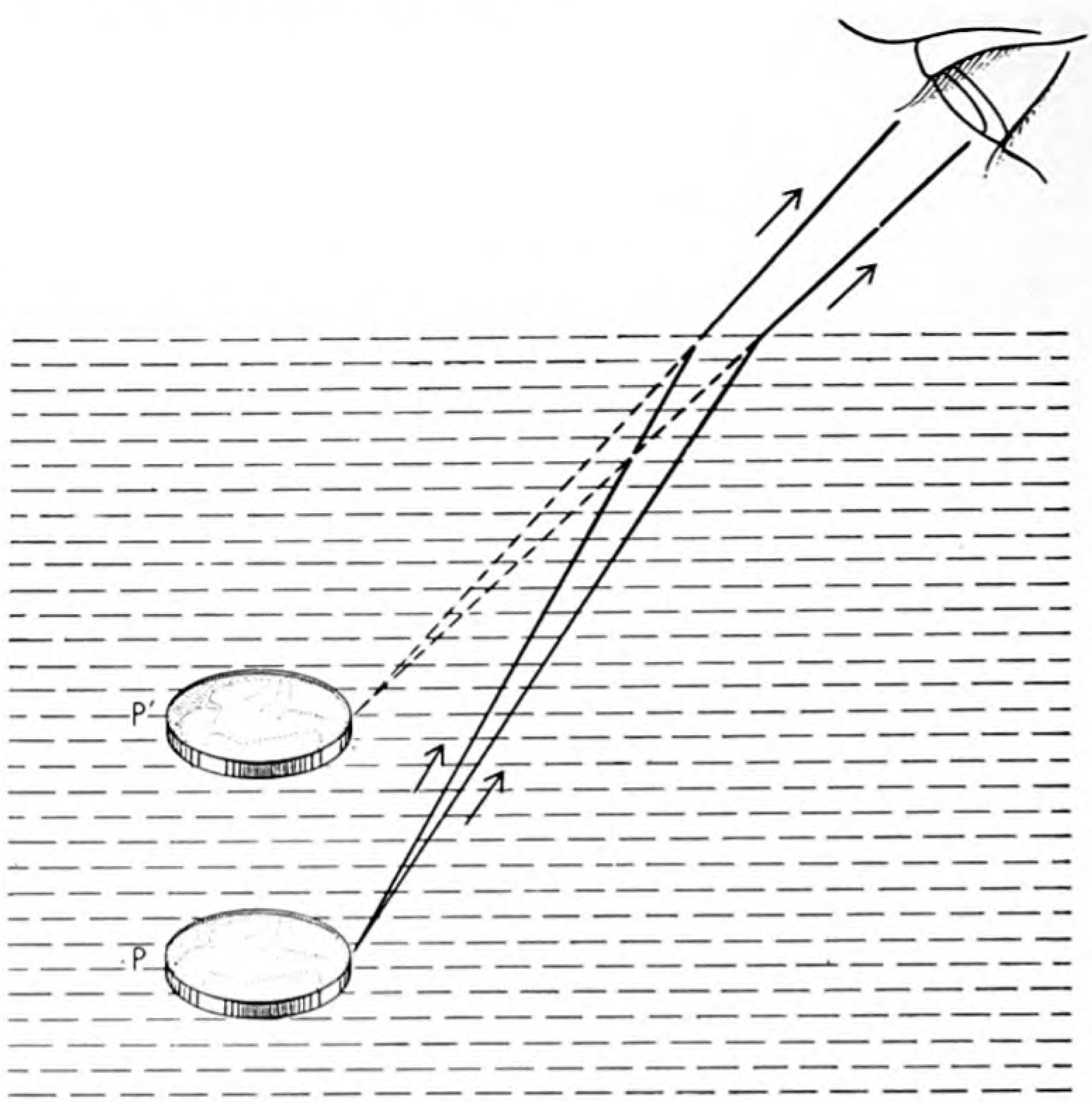
Joule's law is mentioned in many introductory texts but is seldom given as much prominence as Ohm's law. Moreover, it is almost always presented as an entirely independent principle, connected with Ohm's law only through the fact that both involve current and resistance. Actually Joule's law and Ohm's law both express exactly the same thing. Either can be derived from the other or from the principle of the conservation of energy. The electromotive "force" involved in Ohm's law is the energy necessary to move a unit of electrical charge from one end of a conductor to the other. Combining this concept with the fact that the current is merely the rate at which electrical charge is carried through the conductor, we can state Ohm's law as follows: "The energy required to move a unit of electrical charge along the length of any conductor is proportional to the rate at which charge is being moved." For conductors that obey Ohm's law, Joule's law also must hold, because the rate at which energy must be supplied is just the energy per unit charge times the rate at which charge is moved, that is, it is equal to IR times I, or I^2R. According to the principle of conservation of energy, the energy required to move the current against the resistance of the conductor must either be converted to heat or be stored up in some form. When the current in a conductor is steady, there is generally no way that energy can be stored, so the heat given off must be equal to I^2R, in accordance with Joule's law.

It is perfectly true that a discussion of the relation between Ohm's law and Joule's law requires close attention on the part of the student. When he has grasped the reasoning, however, he will have recognized why Ohm's law holds only for steady currents. He will also be able to apply that law to circuits he has never seen before. Most important of all, he will realize that the conservation of energy is a fundamental principle which applies alike to mechanics, to heat and to electricity.

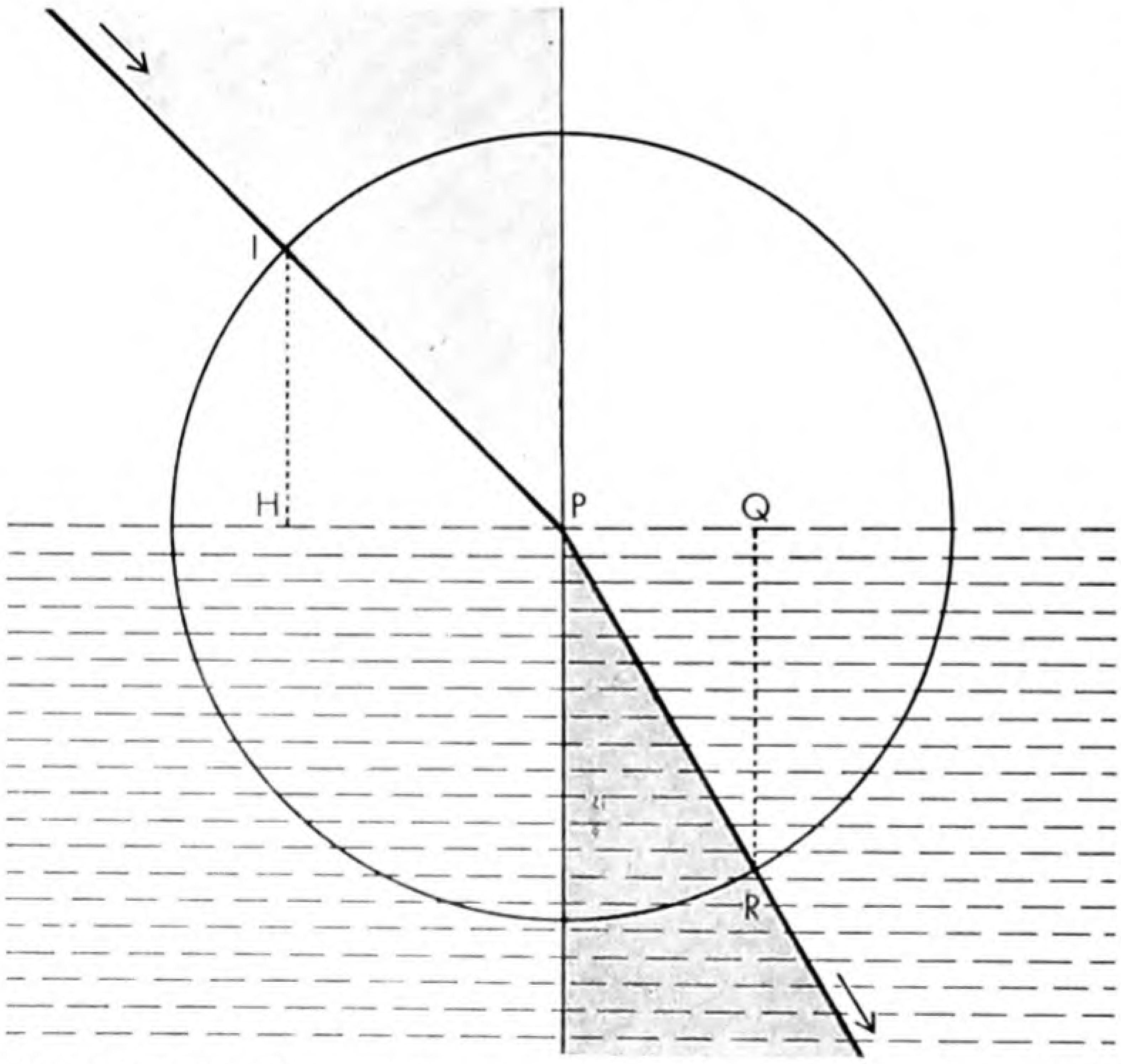
To illustrate further the difference between teaching the principles of physics and merely discussing its application, as courses commonly do, let us take an example in optics. The problem is to explain a virtual image—for instance, the image of an object that you see when you look at it through a magnifying glass. The distinction between a virtual image and a real image (such as is formed by a camera lens) is that the rays of light do not actually pass through the virtual image; rather, the image is a displaced projection formed by extension of straight lines from the eye through the points in the lens where the light from the object itself is bent by refraction [*see upper drawing on next page*]. Now many high-school texts deal with this subject mainly in terms of a "lens equation" which provides a formula for determining the distances of the object and the image and the focal length of the lens. The equation says that the reciprocal of the object's distance from the lens plus the reciprocal of the image's distance is equal to the reciprocal of the focal length, thus: $1/u + 1/v = 1/f$. The student must learn this formula and also a set of rules which say that the distance of the object behind the lens is reckoned as *positive*, but that the distance of the image, if it is behind the lens, is expressed by a *negative* number. He may be given a problem asking what the distance of the image from the lens will be when he looks at an object through a magnifying lens with a focal length of one inch if the lens is held three fourths of an inch from the object. If he remembers the formula and the rules correctly, with a little algebra he can arrive at the answer that v is equal to minus 3—the image is three inches behind the lens. By applying another formula he can learn that the image is four times larger than the object being examined.

The formulas are approximately correct for simple lenses, and they are indeed useful for designers of optical instruments. But they are unlikely to have much meaning for a student or to convey anything about the process by which the image is formed. It is very doubtful that his learning to use the formulas contributes appreciably either to his intellectual development or to his understanding of light or of lenses.

Suppose now that the same subject is taught in a course that emphasizes physical principles rather than their applica-



APPARENT DISPLACEMENT of a coin submerged in water can be understood by tracing paths of light rays from coin to eye. Because of refraction at the surface they appear to be coming from point of broken lines, and this is where coin appears to be.



SNELL'S LAW OF REFRACTION is illustrated by bending of light ray which passes from air to water. Shaded areas show angle of incidence (*top*) and angle of refraction (*bottom*). The index of refraction corresponds to the ratio of the length HP to the length PQ.

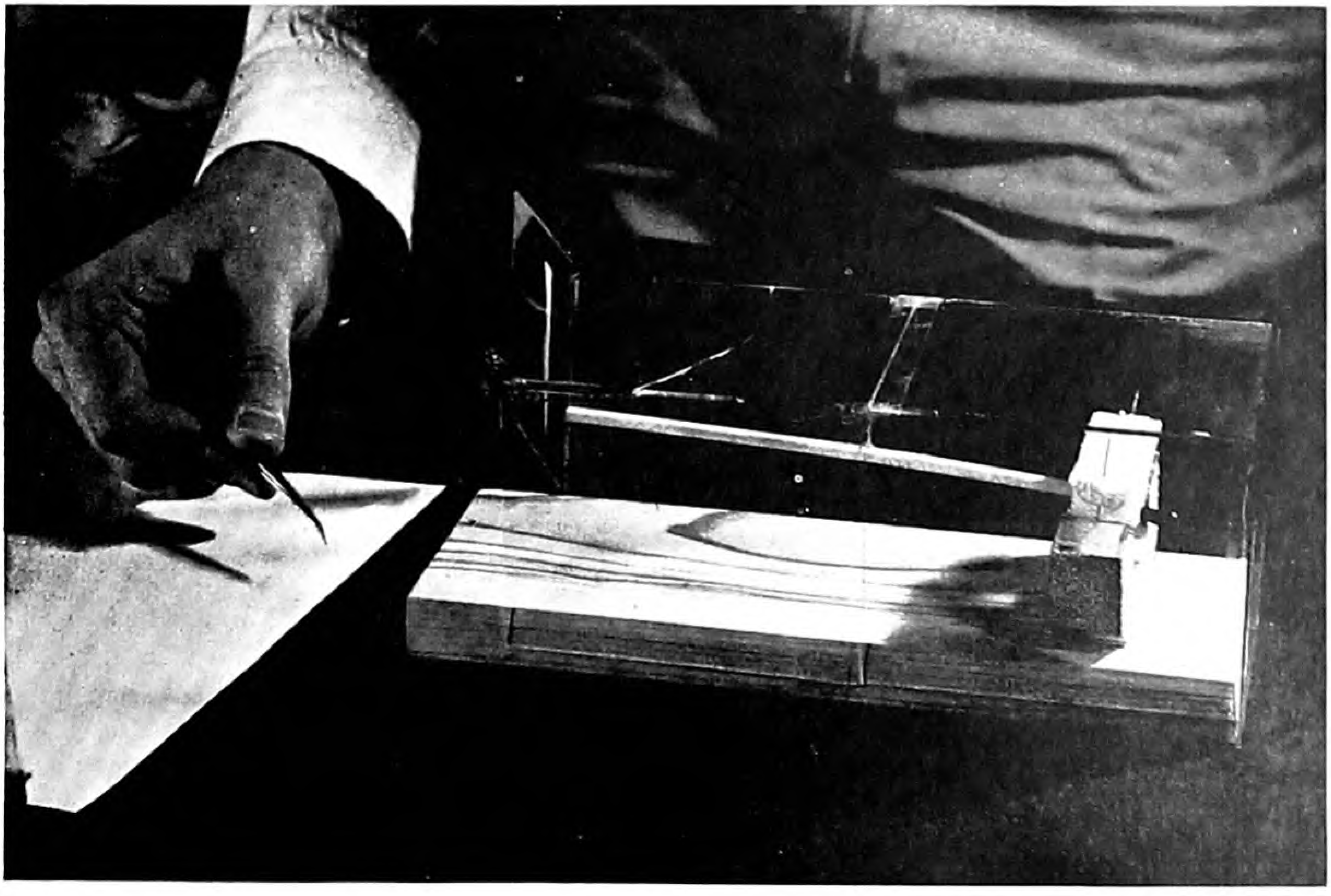
tions. The student may first be asked how we know where an object is. After learning that from each point on the surface of an object rays of light diverge in a cone to the eye (a fact discovered by Franciscus Maurolycus of Messina in the 16th century but seldom emphasized in physics texts), and after studying the law of refraction discovered by Willebrord Snell in 1621, the student may apply that law to a simple problem such as why a coin lying on the bottom of a glass seems to rise when the glass is filled with water. By a simple drawing of the refraction of the light where it passes from the water to the air he can project the displacement of the virtual image, as the illustration at the left shows. Having done this, he can extend the same process to any lens and locate an image by a scale drawing.

This procedure will probably take longer than memorizing and applying the lens equation, but it should give a student an insight into the behavior of light as it passes through a lens, reinforce his knowledge of the law of refraction and equip him with a method that is applicable to all lenses, not just to the thin spherical lenses to which the simple lens equation applies.

New Approaches

These examples are sufficient to illustrate how greatly physics teaching could benefit from a new approach. Even a good teacher is somewhat at the mercy of the current textbooks and of tradition. A poorly prepared one who has been pressed into physics from some other subject has few weapons with which to fight the trend. He cannot be expected, without more time than is allowed by a busy teaching schedule and an overabundance of record-keeping, to learn enough about physics to challenge the text or to present the subject in an enjoyable and exciting way. If physics in the schools is to be improved to the point that the times demand, we must either find a way of attracting more good teachers or we must see that the poorly prepared teacher is given help.

To draw a wider group of students in the secondary schools into physics we must reduce the number of engineering applications taught in the course and devote the time thus gained to the ideas, methods and history of physics. In the past, attempts have been made to interest more students by giving "descriptive" courses, which present physical phenomena but avoid theory and the use of any mathematics beyond arithmetic or a very little algebra. Some of



MICROBALANCE is made of a drinking straw, wood screw and other simple materials. The arm balances on a pair of needles resting on a glass slide. At top left a student places a standard weight in the straw to calibrate the balance. At top right she marks the arm's position. One millimeter on the scale represents 45 micrograms. At bottom a student prepares to weigh a bee's wing.

these courses, particularly when taught with a sense of showmanship, have certainly been popular. No doubt a descriptive course is better than no course at all. But can it give a student the understanding of physical nature and of science that the times require? The distinguishing characteristic of physics is that it involves a central theory, of very broad validity, in terms of which observed phenomena are interpreted and understood. To omit large parts of that theory in order to avoid mathematical difficulties is to deny the able student the intellectual stimulation that is one of the prime rewards of education.

Perhaps good descriptive courses should be developed for the benefit of the lower half of our expanding high-school population, but it is my opinion that we cannot afford such courses for the able students who promise to become the intellectual, political and cultural leaders of the future. For this group there seem to be two possible answers. The first, and best, is to restore solid mathematics courses to the required curriculum for all students in college preparatory courses. The time when we can afford to suppose that intelligent students cannot or will not take a minimum of three years of mathematics in high school is past. The second solution is to include in the physics course whatever mathematics is required for the work, over and above the minimum requirements for graduation from the high school. This would mean that some physics now included in the course would have to be eliminated to make room for the needed mathematics, but we are interested mainly in quality, rather than quantity, in the content of the course. The cost in reduced coverage need not be great if the best secondary-school teachers and the best university physicists can be persuaded to join hands in the discovery of new ways of presenting physics with a high degree of rigor, yet with a minimum of formal mathematics. The history of physics and mathematics supports the amalgamation of the two subjects, for their developments have been so entwined that they are parts of the same intellectual activity. After all, Newton did invent the calculus in order to solve a physical problem.

Training Teachers

The new approaches I have been discussing may sound Utopian, but actually some very promising improvements in the teaching of physics are already under way. Among other things, active steps have been taken in recent years to improve the preparation of teachers. During the past decade a number of colleges and universities have started summer institutes for science teachers. They bring together for six to eight weeks groups of as many as 60 or 70 teachers. The programs include refresher courses, lectures and discussions of current research by active physicists, and laboratory work under better conditions than exist in most schools. The first institutes were financed by a few far-seeing industrial concerns and private foundations; the program is now being supported on a much larger scale by the National Science Foundation, supplying stipends to about 5,000 science teachers each summer, including several hundred in physics. A smaller program, providing teachers with fellowships for a full year of study, has been in operation for two or three years. There is growing support for in-service institutes on campuses near high schools, where teachers study for a few hours each week during the school year. All these training programs are already producing results, and they can be expected, within another five years, to increase greatly the number of well-prepared physics teachers.

To bring physics to many more students in our secondary schools may, however, require more teachers than our educational system and economy can supply. One possible solution would be to enable the good teachers to reach more students by the use of films and television. The Fund for the Advancement of Education made possible a large experiment in this direction. It furnished funds to the schools of the Pittsburgh area for a full-year program of television lectures and demonstrations. The school authorities brought in Harvey White, who had taught physics very successfully to nonscience students at the University of California for many years. He had had experience with television. During the school year of 1956-57 he delivered 162 lectures of 30 minutes each which were carried simultaneously by television to students in 44 schools. At the same time they were filmed. The films are now available through Encyclopaedia Britannica Films and are being used in a number of schools throughout the country. It is perhaps unfortunate that the pace at which this experiment was organized and run did not allow many of the faults of current physics courses to be corrected, but the films should nevertheless be of tremendous help to school systems that cannot find sufficient numbers of teachers who have been adequately trained in physics.

A New Course

A much larger endeavor is being carried out by the Physical Science Study Committee, an organization created by the daring imagination and inspired leadership of Jerrold Zacharias of the Massachusetts Institute of Technology. Supported by the National Science Foundation and a number of private foundations, this committee has brought into its service more than 100 highly qualified individuals, including high-school teachers, research physicists from universities, professional writers, film directors and selected college students. In the past year and a half the committee has considered in detail what might be omitted from the present crowded syllabus to make room for better and more fundamental teaching; it is preparing a radically new textbook; it has made a start on the preparation of new laboratory manuals, teachers' guides and reference materials; it has planned a series of about 70 films; it has designed simple laboratory apparatus, much of which can be built by the students themselves. The course is being developed further by use in a few selected schools, where it has been received with enthusiasm by teachers and students alike. Sufficient materials should be available to allow introduction of the course into many schools by the fall of this year. Five institutes will be conducted this summer to introduce the course to 200 or more teachers.

The work of the Physical Science Study Committee probably represents the most ambitious attack ever made on the problem of presenting any subject to high-school students. It has made use of earlier studies carried out by committees of the American Association of Physics Teachers, the National Science Teachers Association, the American Institute of Physics and others. It has brought together individuals with different backgrounds, including the Nobel laureates Isador I. Rabi and Edward M. Purcell, the Presidential advisers James R. Killian and Vannevar Bush, the writers Mitchell Wilson and Stephen White, the movie producers Frank Capra and Warren Everote, and outstanding teachers such as Judson Cross of Exeter and John Marean of the Reno High School. All those engaged in the enterprise are hopeful that it will produce large results, not only in the teaching of physics but also in elevation of the intellectual level of our schools.

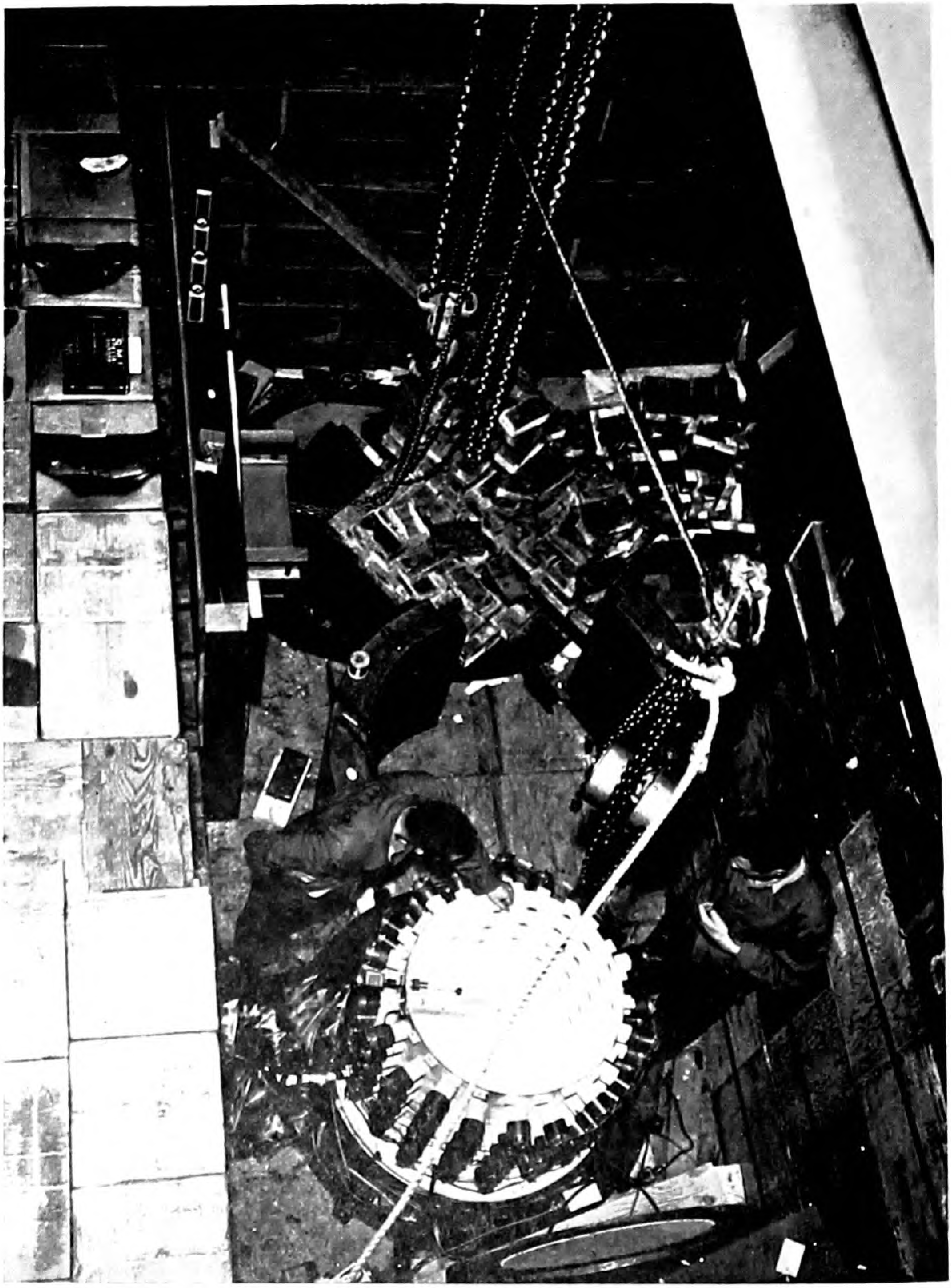
The Author

WALTER C. MICHELS says: "My interest in physics teaching is a natural result of my job. I have been at Bryn Mawr College since 1932, and have been chairman of the Department of Physics since 1936. Bryn Mawr is peculiar in that it has only 650 undergraduate students and yet conducts a full program of graduate work with 200 graduate students, both men and women. We have to give an introductory course in physics that serves at once the future physicist, the major in some other science and the student of liberal arts." One result of Michels's unconventional course was the textbook *Elements of Modern Physics.* Michels is president of the American Association of Physics Teachers. During the last six years he has become increasingly involved in the movement to improve science teaching. He is a graduate of the Rensselaer Polytechnic Institute and the California Institute of Technology.

Bibliography

ENCOURAGING SCIENTIFIC TALENT. Charles C. Cole, Jr. College Entrance Examination Board, 1956.

PHYSICAL SCIENCE STUDY COMMITTEE: A PLANNING CONFERENCE REPORT. *Physics Today*, Vol. 10, No. 3, pages 28-29; March, 1957.



SCINTILLATION COUNTER was used by Frederick Reines, Clyde L. Cowan, Jr., and colleagues in an attempt to detect the neutrino. The counter is the cylindrical object at the bottom. It was set up near a reactor at the Hanford Works of the Atomic Energy Commission. Since this experiment was performed, Reines and Cowan have designed a new experiment using even larger equipment.

SCIENTIFIC AMERICAN January 1956, Vol. 194, No. 1, pp. 58-68 OFFPRINT 230

THE NEUTRINO

by Philip Morrison

For 25 years the theoretical structure of physics has assumed a fundamental particle which has never actually been detected. Its existence may now be confirmed by a remarkable experiment.

The full triumph of classical mechanics came one clear night in the fall of 1846. On that night a German astronomer named Johann Galle pointed the telescope of the Berlin Observatory toward a spot in the heavens where he had been told to look, and there first saw the faint disk of the planet we now call Neptune. His discovery was the most dramatic possible confirmation of Newton's laws of gravitation, and of the calculations of the mathematician Urbain Leverrier, who had predicted from the perturbations in the movement of the planet Uranus that a new planet must lie where Galle found it.

Physics today is looking expectantly for another such discovery. There is a Neptune among its fundamental particles—a strange particle which is on every physicist's list, whose measurements and properties are well known, but which has not yet been "discovered." The particle is called the neutrino. Recognizing that all physical "facts" of nature are no more than inferences, physicists are almost as sure of the existence of the neutrino as they are of anything, but still we will not like to admit that it has really been discovered until it signals its presence by some track or click in our apparatus.

For Christmas not long ago one of the Los Alamos physicists who have been laying ingenious plans to trap the neutrino gave his colleagues a present. Under the gift wrapping they came to a neatly decorated matchbox, clearly labeled: "Guaranteed to contain at least 100 neutrinos." They looked into it warily, and saw a simple empty box. The conceit did not come home until the gift-giver explained it to his friends. The label was literally correct. Such a volume, any such volume on earth, whether a box or your hand or a cupful of water, contains neutrinos. They are moving through it, as they always move, in straight lines at the speed of light.

The neutrinos in us and all around us come from the deep interior of the sun, where they are born in the same nuclear reactions that make the sun shine. Unlike other particles emerging from these reactions, the neutrinos go clean through the sun's great layers, more transparent to them than is the clearest air to the golden terrestrial sunlight, and move out into space. Perhaps 6 or 8 per cent of the total energy released by the sun is in this transcendent form, passing without change or effect through star and space and planet. Even at night neutrinos come streaming to us from the hidden sun, traveling right through the massy earth as though it were not there. The neutrino flux from the sun carries 40,000 times more energy than moonlight, and yet we have never seen, heard or felt its presence.

How such a paradox of a particle has been pressed for decades upon the physicists by hard experience, and how it is today proposed to catch it at last, are the main topics of this article.

The Law of Conservation

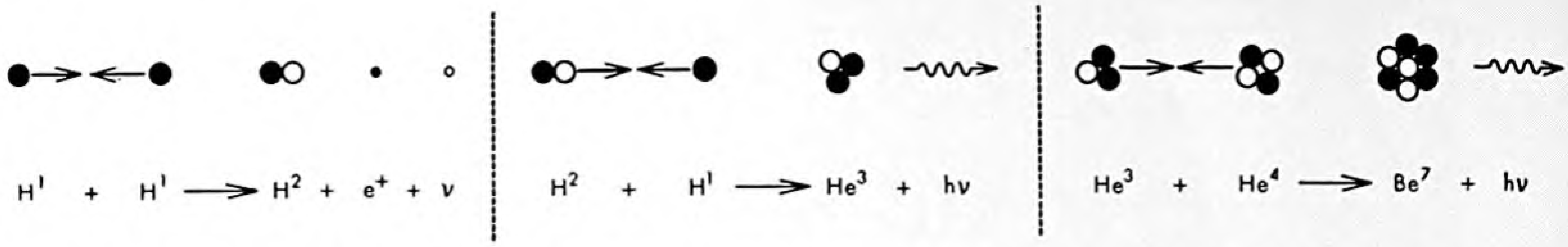
The keystone of physics is the law of conservation of energy. Yet for three decades there has lain within the well-accepted facts of nuclear physics a carefully studied phenomenon which a rigorous and candid observer would have to cite as a *prima facie* contradiction to the famous law. This scandal lies in a strange disappearance of energy when a neutron sheds an electron—the phenomenon known as beta-decay.

A wide variety of evidence indicates that the uncharged neutron, like its charged fellows the proton and the electron, has a perfectly definite mass. Now the neutron, building block though it is, is not fully stable. It spontaneously decays into a proton and an electron. The neutron's half-life is some 12 minutes, which is to say that if 1,000 neutrons are kept free from any interaction with the outside world, after 12 minutes only 500 will remain neutrons, and after about 24 minutes, only 250.

By the principle of the convertibility of mass into energy, the total energy content of a stationary neutron is just the energy equivalent of its mass. Now the products of its decay—proton and electron—do not add up in mass to the initial mass of the neutron. The missing mass has been converted into energy, and this appears as kinetic energy of the two product particles, which fly apart after the neutron's disintegration.

The law of the conservation of energy says that the kinetic energy shared by the neutron's decay products (the lion's share goes to the light electron) must be precisely equivalent to the lost mass. Their kinetic energy has been measured, and its maximum value has been determined to be about 780,000 electron-volts. But the numerous measurements have shown that comparatively seldom does the kinetic energy released by a neutron's decay reach this value. Sometimes the proton plus electron have a total kinetic energy of only a few thousand electron-volts. The decay energies observed range over the whole spectrum from zero to 780,000 electron-volts. On the average, only a fraction of the maximum expected energy release appears when a neutron (or any radioactive atomic nucleus) splits by beta-decay.

What happens to the missing energy? Decades ago nuclear physicists came to the plausible conclusion that it must go off with some undetected particle. They



PROTON-PROTON SEQUENÇE is one of two nuclear processes in the sun which are assumed to shower the earth with neutrinos. In the first reaction of the sequence two protons (*nuclei of hydrogen 1, or* H^1) collide to yield a deuteron (*nucleus of hydrogen 2, or* H^2), a positive electron (e^+) and a neutrino (ν). In the second reaction the deuteron collides with a proton to yield a triton (*nucleus of hydrogen 3, or* H^3) and a gamma ray ($h\nu$). In the third reaction the triton collides with an alpha particle (*nucleus of helium*

tried hard to catch the invisible particle in their best counters, and in a foot or two of lead. But they failed. And in a dark year there came the thought that perhaps here in the fundamental processes of the nucleus the conservation of energy actually failed.

After the successes of the quantum theory, which no less than classical physics is based upon energy conservation, Wolfgang Pauli first sketched and Enrico Fermi later worked out in detail the properties of the presumed particle that must fly off in beta-decay if energy was to be conserved. It was easy to see what some of these properties must be.

First, the particle must have no charge, for balancing charges (the positive proton and negative electron) come out of the decaying neutron. Secondly, the particle must be practically weightless, according to careful measurements of the energy and mass relations of the various particles involved. For simplicity it is considered to have zero mass. Thirdly, the particle must have angular momentum, *i.e.*, spin. The neutron's angular momentum is one half a quantum unit; when it decays it produces two particles each with one-half unit spin; so to conserve angular momentum the invisible particle should have one-half unit spin.

Such a particle, with no charge, no mass, spin one half and birth only in beta-decay, was postulated by Pauli and by Fermi. It became known as the neutrino—"little neutral one"—in the beautiful language of Fermi.

But now we are on thin ice. Faced with a failure of energy conservation, physicists refuse to admit it but instead postulate an unseen and perhaps unseeable particle—a little neutral one so cunningly designed that it has no properties other than those which will preserve the laws of conservation. How does this differ from plain failure of the conservation laws?

The Challenge

The first attempt at an answer was to examine carefully the scene of the crime. The particles issuing from a decaying neutron or atomic nucleus carry momentum as well as energy; hence they should recoil from each other, just as the firing of a shell causes the cannon to recoil. If only a proton and an electron came off, conservation of momentum would require them to move in opposite directions along one straight line. But if a neutrino carrying momentum also comes out, the other pair must recoil from *it*, and their tracks should form a V. A number of elegant experiments recording the beta-decay of various nuclei and of the neutron itself have proved that this is indeed the form that the tracks actually take.

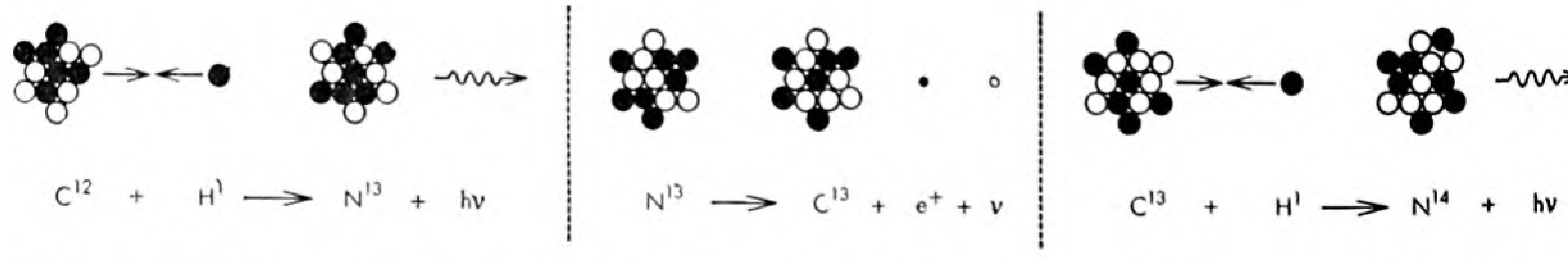
The critics can still say: Momentum and energy are intimately connected, so your recoil experiment proves nothing except that both energy and momentum are lost in a single parcel. Until you can somehow trace the missing energy, momentum and the rest, you are merely balancing the books with a fictitious entry.

There is only one sure answer to the criticism. The missing energy, the "little neutral one," must be caught. The energy it carries must be reconverted into some measurable form, or the neutrino, however universally accepted, remains but a sign of the physicist's ignorance.

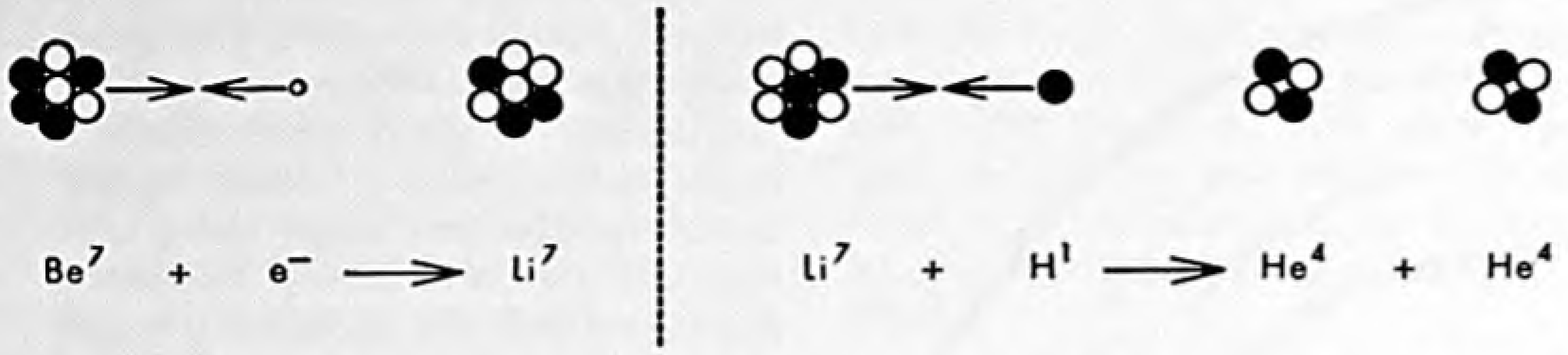
Of course the neutrino theory has by no means been unfruitful. It has provided a basis for interpreting various measurable features of beta-decay, for predicting the approximate lifetimes of all beta-decaying nuclei, and so on. We could not give up the real triumphs of the neutrino postulate without tearing the present closely webbed fabric of nuclear physics. But still the challenge to track down the neutrino itself remains.

The behavior of physicists in the face of this challenge may seem unworthy. For a generation we have been satisfied to use the neutrino theory and explore its manifold ramifications, but slow to respond to the challenge to deliver up the thing itself as a legitimate member of the particle family, not a mere bookkeeper's evasion. Why have we been so laggard in seeking out the neutrino directly?

The answer is simple, and yet in some ways wondrous. Compared to other nu-



CARBON CYCLE is the second solar nuclear process which may produce neutrinos. In the first reaction of the cycle carbon 12 (C^{12}) collides with a proton to yield nitrogen 13 (N^{13}) and a gamma ray. In the second reaction nitrogen 13 decays spontaneously into carbon 13 (C^{13}), a positive electron and a neutrino. In the third reaction carbon 13 collides with a proton to yield nitrogen 14 (N^{14})



4, *or* He⁴) to yield beryllium 7 (Be⁷) and a gamma ray. In the fourth reaction the beryllium 7 collides with a negative electron (e⁻) to yield lithium 7 (Li⁷). In the fifth and final reaction of the sequence lithium 7 collides with a proton to yield two alpha particles. The protons are represented by the larger black balls; the neutrons, by white balls of the same size.

clear events, beta-decay is fantastically slow, and therefore rare. It takes a neutron some 18 minutes on the average to emit an electron; with the same energy available, a gamma ray would come out in a millionth of a billionth of a second. The prodigiously slow beta process has a far smaller probability than any other method of nuclear decay. Beta-decay is much rarer, on the nuclear time scale, than death by meteorite bombardment among men. It is only the chance that some nuclei are immortal except for beta-decay that allows us to observe this event at all.

Now this slowness of decay implies that the opposite process, the capture of a passing neutrino by a nucleus, also is slow and rare. Gamma rays, which are notoriously unlikely to interact with nuclei, will travel on the average through eight or 10 feet of lead before they do so. But a neutrino, to interact with a nucleus, must travel on the average through about 50 light-years of solid lead! A shielding wall capable of thinning out a beam of neutrinos would have to be as thick as a hundred million stars. To all intents and purposes neutrinos simply do not see solid matter at all. Here is the nub of the difficulty. The neutrino is almost uncapturable.

The Challenge Accepted

Deterred by the logic of the matter, physicists did not begin to think seriously about hunting neutrinos until great masses of radioactive material became available in the fission products of uranium reactors. With a truly prodigal effort, they now seek to catch the almost uncatchable neutrino. From the high-power reactors stream currents of neutrinos which rival the sun's beam. They pass through the shielding walls and fly in perfectly undeviating straight lines from the place of their birth to outer space. Patrolling these great beams near their source, physicists hope to capture an occasional unlucky neutrino. Last month a group of workers began the vigil by placing their neutrino detectors in the beam from the most powerful reactors in the world—those at the Savannah River plant of the Atomic Energy Commission.

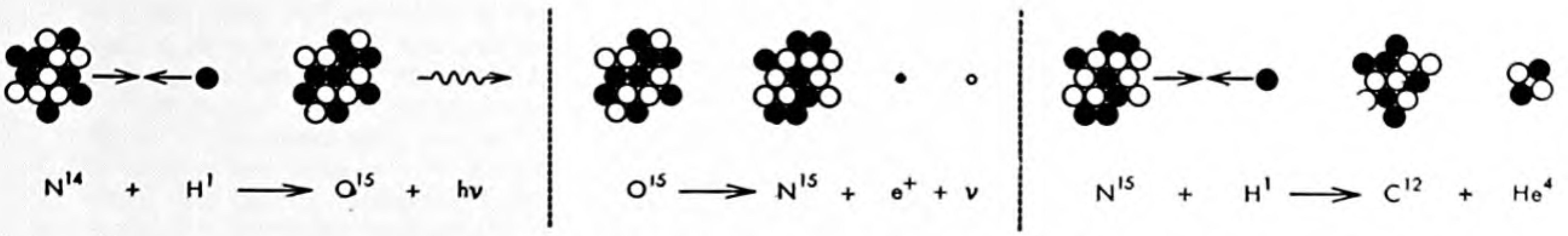
The more spectacular of two schemes for trapping the little neutral ones is based on the technique of scintillation counting. Tireless photocells will keep a continuous surveillance over a large tank full of a clear scintillating liquid, recording the flashes of light that signal ionizations among the nuclei in the liquid. A Geiger counter can record only the ionizing events occurring in a few milligrams of gas; the scintillation tank scheme makes it possible to patrol tons of nuclei at a time.

But the capture of the neutrino is a rare event indeed, and when one looks for the highly improbable, he must be prepared to see many other events, probable and improbable, which are irrelevant to what he is looking for. The tons of scintillator liquid will flash many times—from traces of radioactive dirt within the liquid, from cosmic rays, from escaping particles other than neutrinos which may come from the reactor. It is not enough simply to look for flashes; it is necessary to discriminate between those which arise from neutrino capture and those which arise from various other rare but much less interesting causes.

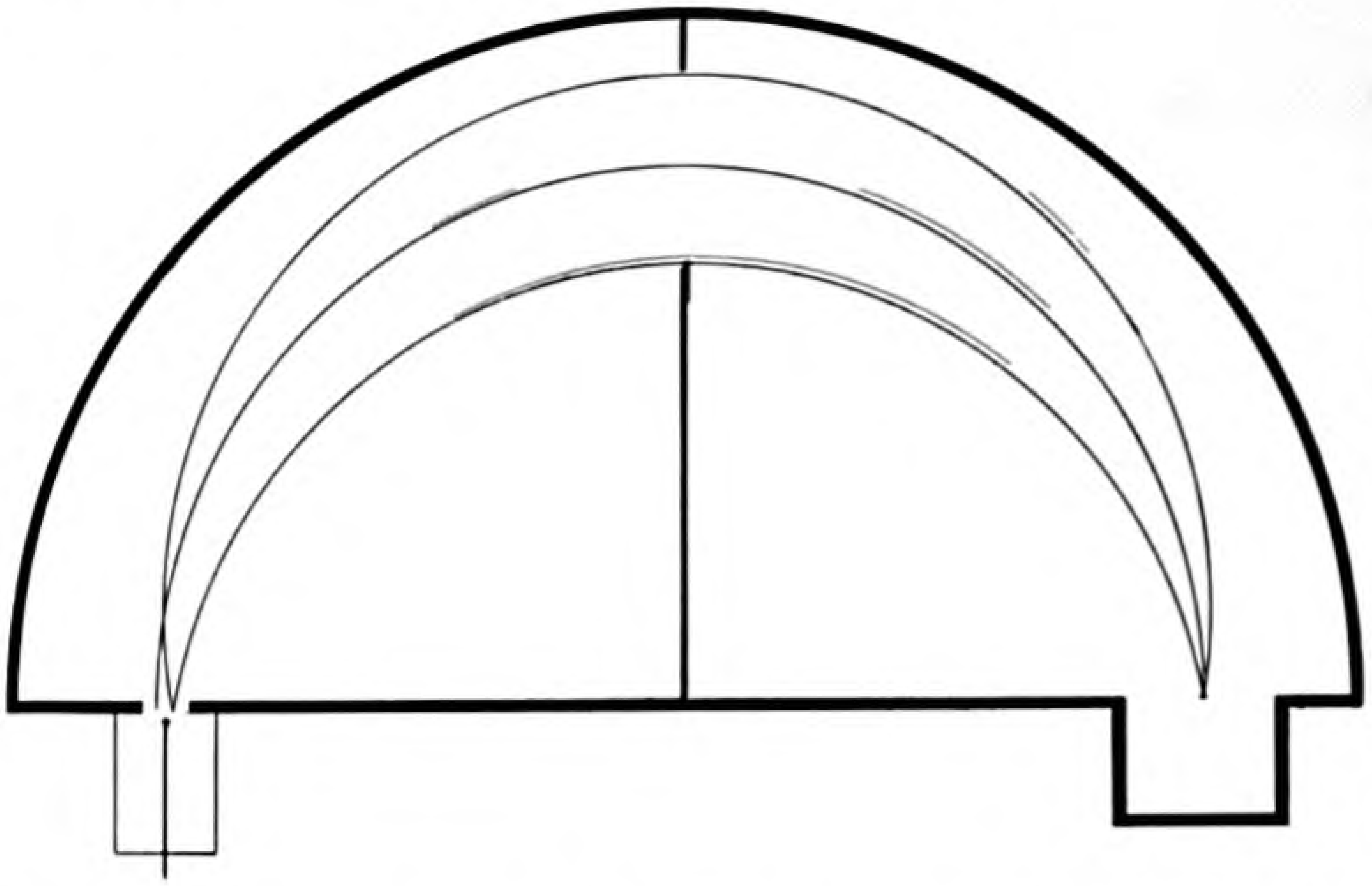
A group of Los Alamos investigators led by Clyde L. Cowan, Jr., Francis B. Harrison and Frederick Reines invented, and is pursuing with true virtuosity, an ingenious scheme to perform this feat. The reasoning is as follows. Capture of a neutrino being the reverse of its emission, the precise opposite of beta-decay must occur: that is to say, a proton plus an electron plus a neutrino plus energy combine to form a neutron. Now there is another type of "beta-decay" reaction in which, instead of a negative electron being absorbed, a positive electron, or positron, is emitted. Electric charge is conserved equally well in either case: in the second case the positive proton becomes a neutral particle not by absorbing a balancing negative charge but simply by releasing its own positive charge in the form of a positron. That such reactions actually occur has been verified by experiments. And so we can rewrite the equation for neutrino capture to say that a proton plus a neutrino plus energy may combine to yield a neutron and a positron. The ingenious Los Alamos group proposes to detect this event by observing its products—the neutron and the positron.

A Subtle Detection

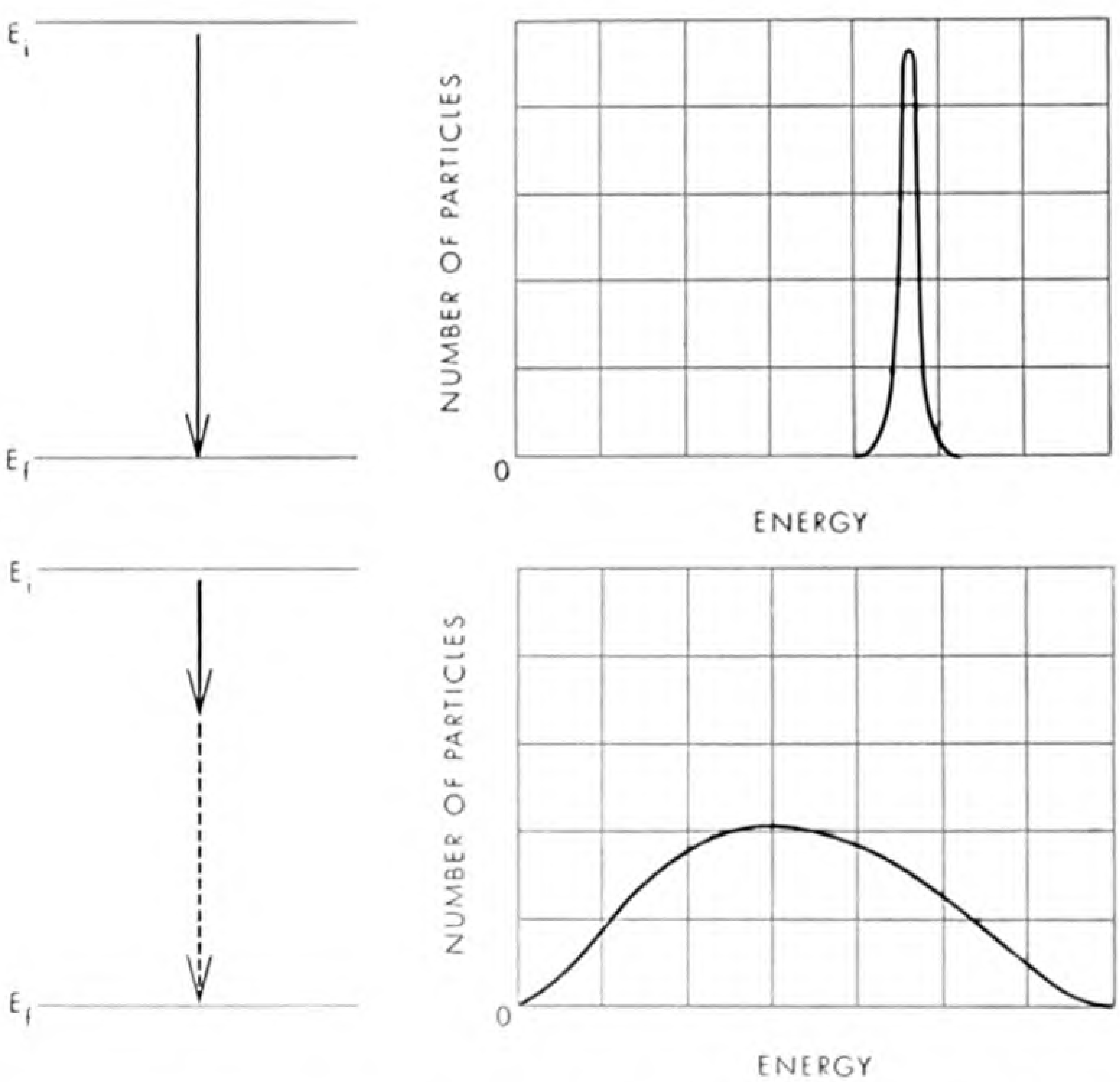
The energy necessary to convert a proton into neutron and positron is supplied by the incoming neutrino. Neutrinos emerging from the fission products in a reactor are estimated to have, as a rule, something of the order of one mil-



and a gamma ray. In the fourth reaction nitrogen 14 collides with a proton to yield oxygen 15 (O¹⁵) and a gamma ray. In the fifth reaction oxygen 15 decays into nitrogen 15 (N¹⁵), a positive electron and a neutrino. In the sixth reaction nitrogen 15 collides with a proton to yield carbon 12 and an alpha particle. The cycle makes an alpha particle, two electrons, three gamma rays and two neutrinos.



SPECTROMETER is used to measure the energies of beta particles. The apparatus in this diagram is placed in a magnetic field, in which the lines of force are perpendicular to the page. The particles emitted by the small sample at right describe curved paths and are counted at left. Particles of the same energy come to a focus at the same point. Thus the counter may be used to scan the various particle energies by changing the strength of the magnetic field.



ALPHA- AND BETA-DECAY are accompanied by changes in the energy level of the nucleus. The diagram at upper left depicts the energy levels of a typical alpha-emitting isotope before and after decay. The curve at upper right shows how the energies of particles from this isotope are distributed. The fact that the particles occupy a narrow range of energies indicates that the difference between the initial energy of a single nucleus (E_i) and the final energy (E_f) is always accounted for by the energy of an alpha particle (*arrow*). The diagram at lower left depicts the energy levels of a typical beta-emitting isotope before and after decay. The curve at lower right shows how the energies of particles from this isotope are distributed. The fact that the particles occupy a broad range of energies indicates that the difference between the initial energy and the final energy is seldom accounted for by a beta particle (*solid arrow*). The missing energy is presumably carried off by a neutrino.

lion electron-volts more energy than is required for this conversion. The excess energy goes off as kinetic energy of the product neutron and positron—most of it in the positron, since it is much lighter. The charged positron ionizes atoms in its path as it goes, and thereby produces a good-sized flash in a scintillating liquid. After it has traveled a centimeter or two, which takes no more than a hundredth of a billionth of a second or so, the positron comes to rest, having spent all its kinetic energy.

The flash from ionization by the moving positron is the first visible sign of the neutrino's capture, but it is by no means the last: there is more to come. The moment it comes to rest, the positron combines with an electron (there are plenty handy). When it does, the mass of the pair is instantly annihilated and turned into energy. The energy flies off as two gamma rays, traveling in opposite directions from the source and each amounting to about one half million electron-volts. After moving a foot or two, each gamma ray gives rise to an ionizing electron which produces a flash in the counter. So every neutrino capture that gives birth to an energetic positron should be signaled by three virtually simultaneous flashes in the scintillating liquid—a flash from ionization by the moving positron and a pair of flashes in different spots from the gamma rays.

Meanwhile the newborn neutron itself has moved off slowly, with very little kinetic energy. There is no flash to mark its passage, for the neutral particle cannot ionize. It wanders about the liquid in its usual random walk, slowing down to thermal motion as it goes. Eventually it is captured by some nucleus. Now the canny experimenter may add to his liquid scintillator some cadmium, which has a pronounced affinity for slowed neutrons. The cadmium nucleus seizes a neutron so vigorously that a few gamma rays are emitted in the process. In this moment of capture, therefore, the neutron signals its presence by a flash, as the positron did earlier. The neutron signal comes after a considerable delay, because the particle has traveled a yard or two, at a relatively slow speed, before its capture. This delay is some 10 or more microseconds, and it can be measured accurately by electronic techniques.

The plan of the experiment at the Savannah River plant is this. A layer of liquid containing protons and doped with cadmium is placed like a sandwich filling between two thick scintillating layers sensitive to gamma rays. Photomultipliers watch all carefully. When a

neutrino is captured by a proton, the resulting birth of a positron is instantly signaled by a flash in the sandwich layer. Practically simultaneously, after a time too short to measure, there comes a flash in each of the two "bread" layers, produced by the two gamma rays from the annihilation of the positron. A few millionths of a second later the capture of the wandering neutron by cadmium releases gamma rays whose flashes are seen in all three layers. In short, every neutrino capture is marked by two sets of flashes in all three layers, one following the other after a precisely stipulated interval. Moreover, the energy of the positron-annihilation flash helps to identify the event: it should total about one million electron-volts.

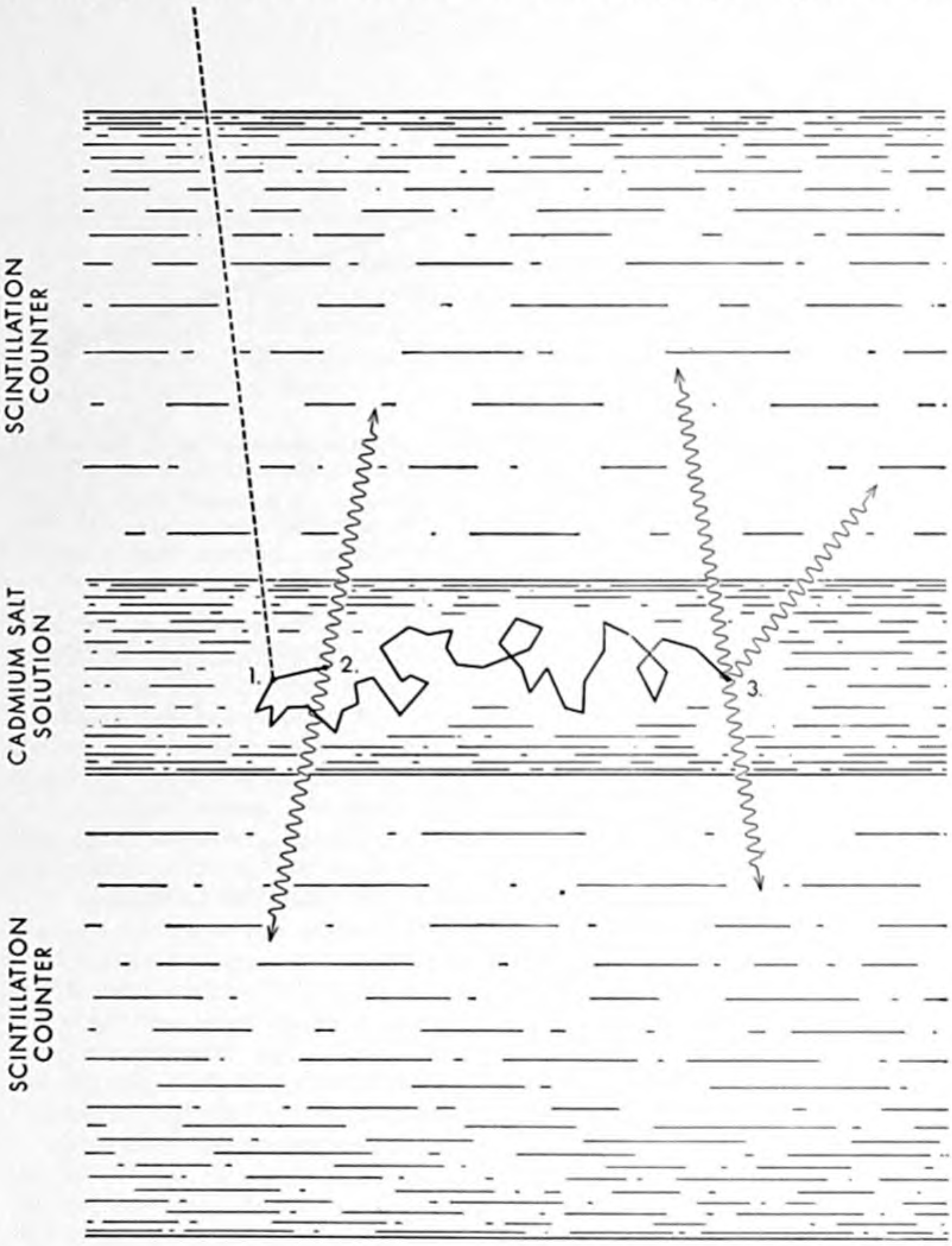
Combinations of events which simulate this pattern may occur by accident in the counter, but they are too infrequent to cause real trouble. Their spuriousness can, and will, be established by control experiments. Cosmic ray particles can cause spurious events (and did so seriously in the earliest versions of this experiment). These fast charged particles may trigger all three layers, and in addition liberate a neutron which will give a delayed pulse. But the energy they give to each layer is large, and the first flash will be too bright, and give away the spurious character of the event.

The whole apparatus is of prodigious size. Where most experiments of the kind use half a dozen photomultiplier tubes and their associated amplifiers, the neutrino searchers use 500. Where scintillation counters are normally counted big if they use a few gallons of liquid, this experiment uses 10 or 12 tons. The needs of the project have led to a whole complex of ingenious and painstaking developments. The chemical firm producing the scintillating liquid, which used to make it in quart amounts, has been persuaded to manufacture and purify it by the ton. A special tank truck has been built to transport the precious fluid in an inert atmosphere from the factory to the scene of the experiment; it must be kept minutely clean and oxygen-free throughout the long journey. The tank where the experiments are performed must be lined with a special glossy-white coating, to lose next to no light at all. A chemical must be added to the scintillating liquid to give the flashes a color which will be reflected most efficiently by the gleaming tank walls. The flashes that carry all the information for which the neutrino hunters are searching are too faint to be seen by the naked eye, and no effort must be spared to make sure that all of them are detected by the sensitive photomultipliers. Hundreds of the latter are required, and whole banks of other electronic gear. The detector tank itself is encased in lead and buried deep in the building housing the great Savannah River reactor.

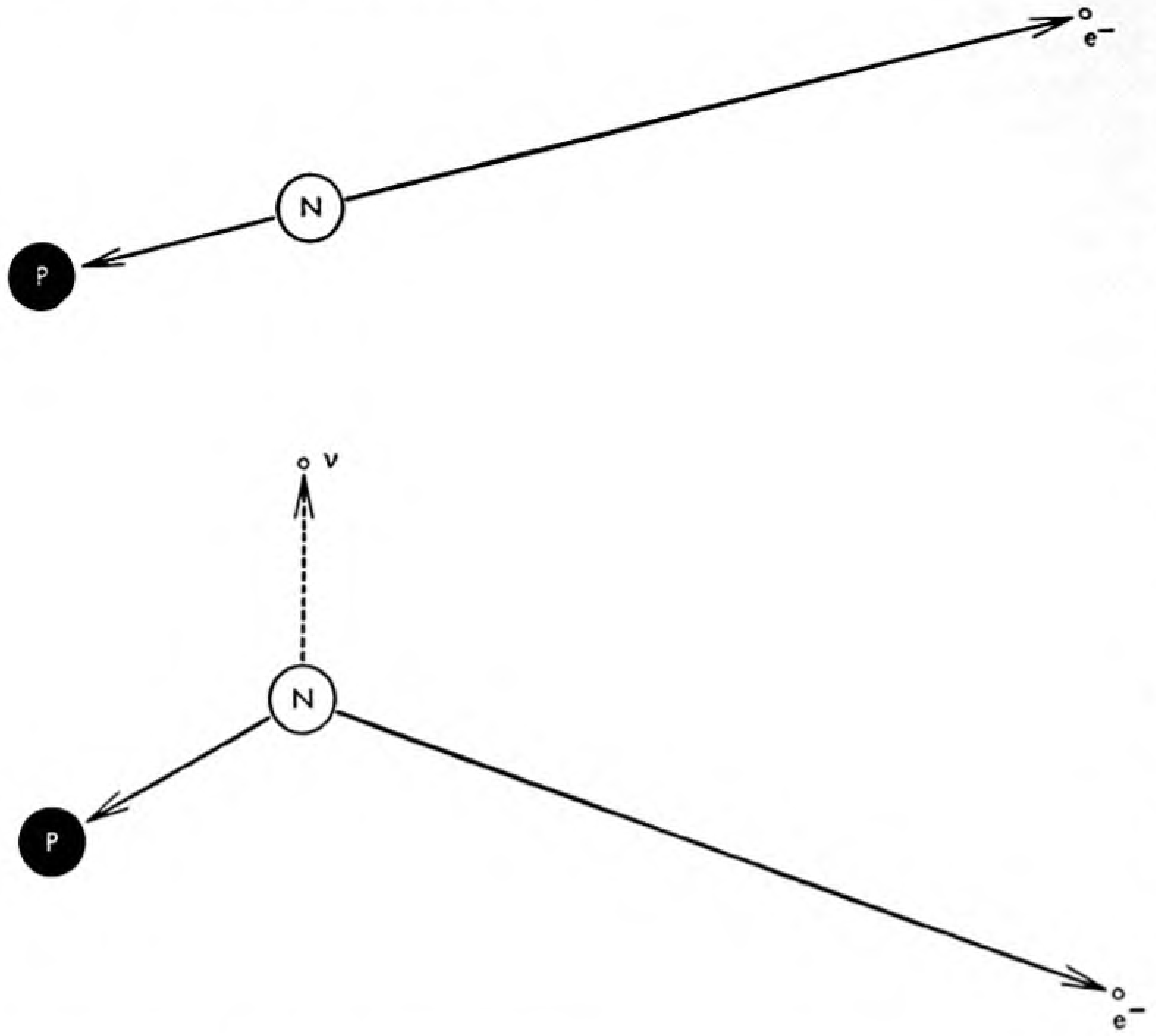
After building and testing all this equipment, the experimenters began to install it at Savannah River last month. There they will count patiently, hour after hour, waiting for the evidence which they hope will restore the conservation of energy to honesty and will help them play the Galle to Fermi's Leverrier in the physics of the 20th century.

The Search by Chemists

Meanwhile another group pursues the search by a very different method. Just as the scintillation-counter plan is a *tour de force* of modern physical technique, so there is a related piece of consummate skill which belongs to the chemist. The trick of identifying the capture of neu-



DETECTION OF THE NEUTRINO is planned by Cowan and Reines on the basis of the events outlined in this diagram. A neutrino (*dotted line*) encounters a proton (1), causing it to change into a neutron (*zigzag path*) and a positive electron. In passing through the cadmium salt solution the positive electron causes a pulse of ionization. Then it is annihilated in an encounter with a negative electron (2). This gives rise to two gamma rays (*wavy lines*), one of which causes a pulse in the top scintillation counter and the other a pulse in the bottom scintillation counter. The neutron wanders for several microseconds until it is absorbed by a cadmium nucleus (3). This gives rise to three gamma rays. Thus when two consecutive pulses, separated by an interval of several microseconds, occur in both the cadmium solution and the scintillation counters, a neutrino is assumed to have been detected.



RADIOACTIVE DECAY OF THE NEUTRON also requires the neutrino. If the neutron (N) simply decayed into a proton (P) and a negative electron (e^-), the conservation of linear momentum would require that the decay particles go off in exactly opposite directions (*top drawing*). Actually they go off at an angle to each other (*bottom drawing*). This indicates that the missing linear momentum is carried off by a neutrino (*broken arrow*).

trinos by detecting products of the reaction may, it is hoped, be performed by chemical means. In this case the targets are not protons but nuclei of chlorine atoms. The isotope chlorine 37 may capture a neutrino and turn into an atom of the rare gas argon, emitting a negative electron in the process. The atom so produced, argon 37, is itself radioactive.

Its creation may be detected by a simple plan. A large tankful of a liquid containing chlorine is swept clean of argon (certainly of any radioactive argon) by a stream of helium bubbled through it, and then the liquid is exposed to neutrinos from a reactor. After some days' exposure, the tank again is swept with helium gas to bring out any argon that may have been formed. The argon (of which there must always be at least a slight trace) is separated from the helium by well-known techniques of physical chemistry, and a Geiger counter is used to see if there is any radioactive argon in the sample.

This method is based on such a refined chemical search for a few atoms in a great vat of solvent that it is surely the archetype of all needle-in-the-haystack endeavors. There are against it two objections. The first, of course, is that it leaves open the question of just what did make the active argon. Careful controls might dispose of this question. But then there is left a more fundamental, if not more serious, objection flowing from the theory of fundamental particles. If we reverse the neutrino-capture reaction, the theory says that the decay products must be chlorine 37, a neutrino and a positron rather than an electron. This means that the neutrino is not precisely the same as in a normal beta-decay. Consequently the beta-decay neutrinos sent out by a reactor may not transmute chlorine into argon. However, it is possible that the two kinds of neutrino are actually equivalent in every respect, so that the experiment will really work.

Raymond Davis, Jr., of the Brookhaven National Laboratory will measure the argon activity induced in some four tons of carbon tetrachloride in a tank at the Savannah pile. It is a happy circumstance that both the scintillation counter and the chlorine experiments are being carried on at the same time and the same place, for the two should complement each other.

It would be ungracious not to mention that these are not the first experiments to seek out the neutrino directly. Quite a few brave experimenters have tried before, but on too small a scale. The chlorine experiment was first planned at Chalk River in Canada by Bruno Pontecorvo, who may for all we know now be completing it near some big reactor in the U.S.S.R. The proton-capture experiment in a smaller version was tried by the Los Alamos group at Hanford last year, and a doubtful positive result obtained. Most physicists will be willing to base conclusions only upon the really powerful effort now under way.

A New Astronomy

The neutrino provides in principle a new kind of astronomy. Until now the only radiations from outer space that we have studied are visible light and microwave radio. The neutrino beam from the sun and stars also comes to us bearing information about the universe, and we shall surely some day read a part of that text. The chlorine-capture experiment, with an increase of a factor of a hundred or so in bulk, and a well-shielded mine or the deep sea to work in, might measure the sun's neutrino flux. Modifications in the scintillation scheme also are under consideration. It is not too much to hope that some day we may directly verify the nuclear reactions in the sun's center by a study of their neutrino emission.

But suppose the experiments do not work? Suppose no neutrino counts are seen? The logical chain is pretty tight; the defeat would mean to many that energy conservation had at last really failed us, or, almost as bad for our theories, that a reaction was not accompanied by its inverse among the fundamental particles. We should be loath to accept either of these conclusions. So far I have thought of only one way out—a desperate evasion. The neutrino might leave the scene of its birth all right, starting life as a perfectly proper Fermi neutrino. But it might be unstable and decay into three other neutrinos (two is forbidden for dynamical reasons), which could of course not all be simultaneously caught at the other end to invert the reaction, because their directions would slightly diverge. This possible behavior would leave the beta-decay theory intact. But it would make the neutrino even more evanescent a notion than it now is, and would only pass the trouble on to later generations of physicists. It will be far better if the patient experimenters are successful, and if their scintillator clearly displays those few oscilloscope traces each hour which will mean that the fugitive neutrino has been caught at last.

The Author

PHILIP MORRISON is associate professor of physics at Cornell University, where he has done pioneer work applying physical theory in a number of fields, including microbiology. He graduated from the Carnegie Institute of Technology in 1936, then studied theoretical physics under J. Robert Oppenheimer at the University of California, where he received his doctorate in 1940. When World War II broke out, Morrison left a lectureship at the University of Illinois to join the Metallurgical Laboratory of the University of Chicago, and later became a group leader at the Los Alamos Laboratory of the Manhattan District. Morrison joined the faculty at Cornell in 1946.

Bibliography

DETECTION OF THE FREE NEUTRINO. F. Reines and C. L. Cowan, Jr., in *The Physical Review*, Vol. 92, No. 3, pages 830-831; November 1, 1953.

INTRODUCTORY NUCLEAR PHYSICS. David Halliday. John Wiley & Sons, Inc., 1955.

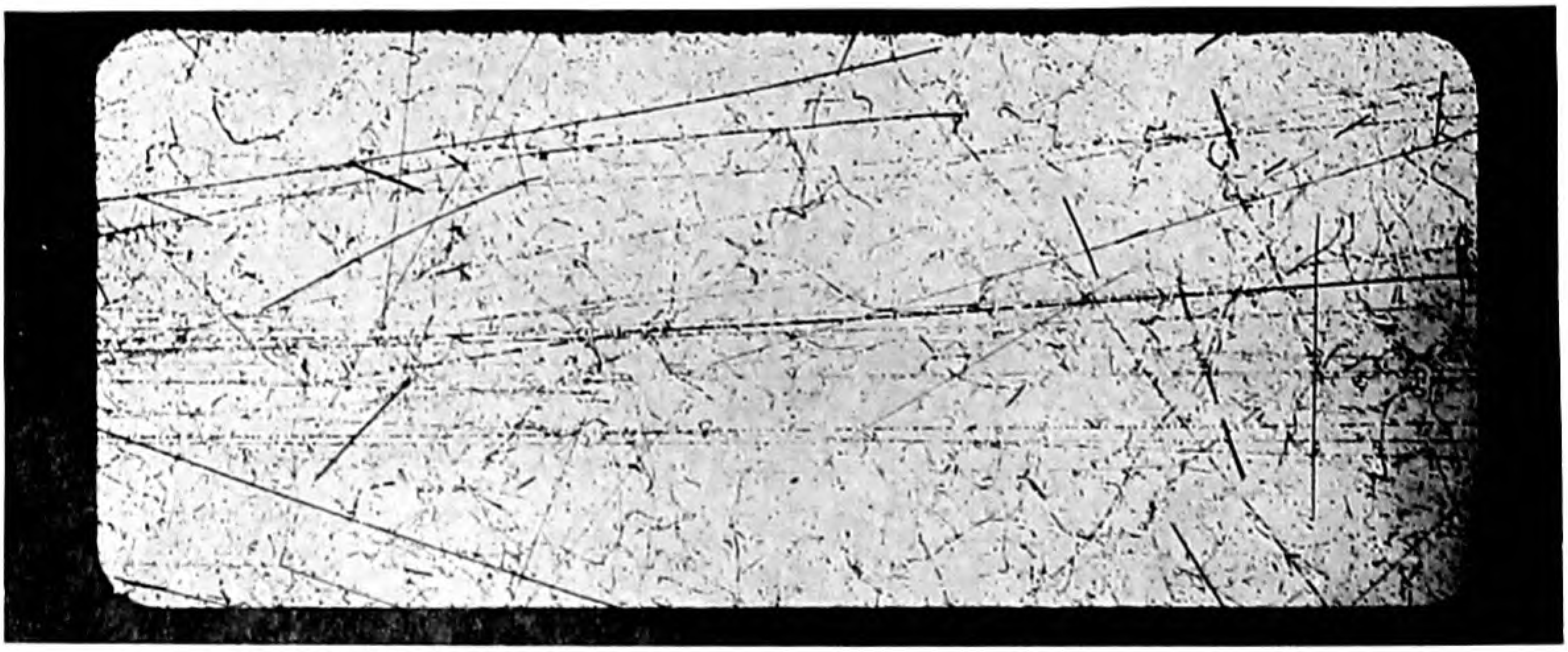
SCIENTIFIC AMERICAN April 1957, Vol. 196, No. 4, pp. 45–53 OFFPRINT 231

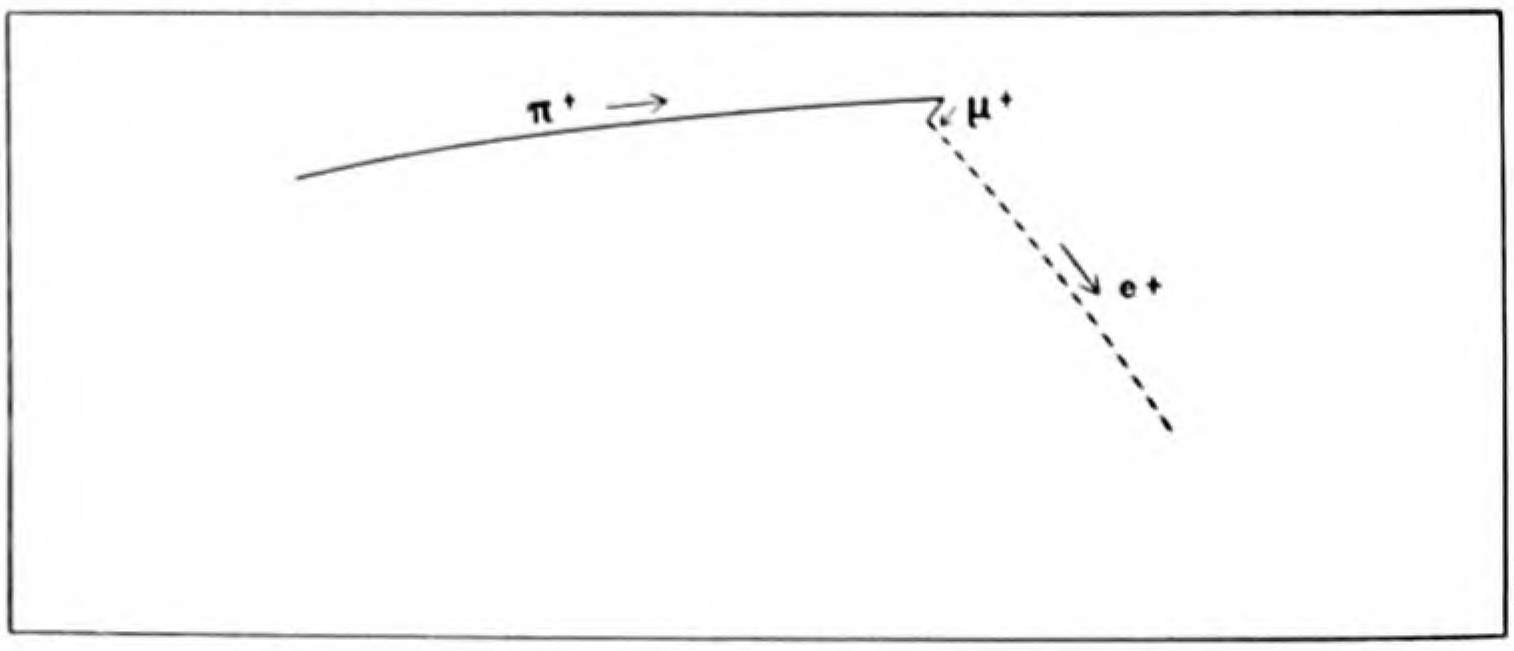
THE OVERTHROW OF PARITY

by Philip Morrison

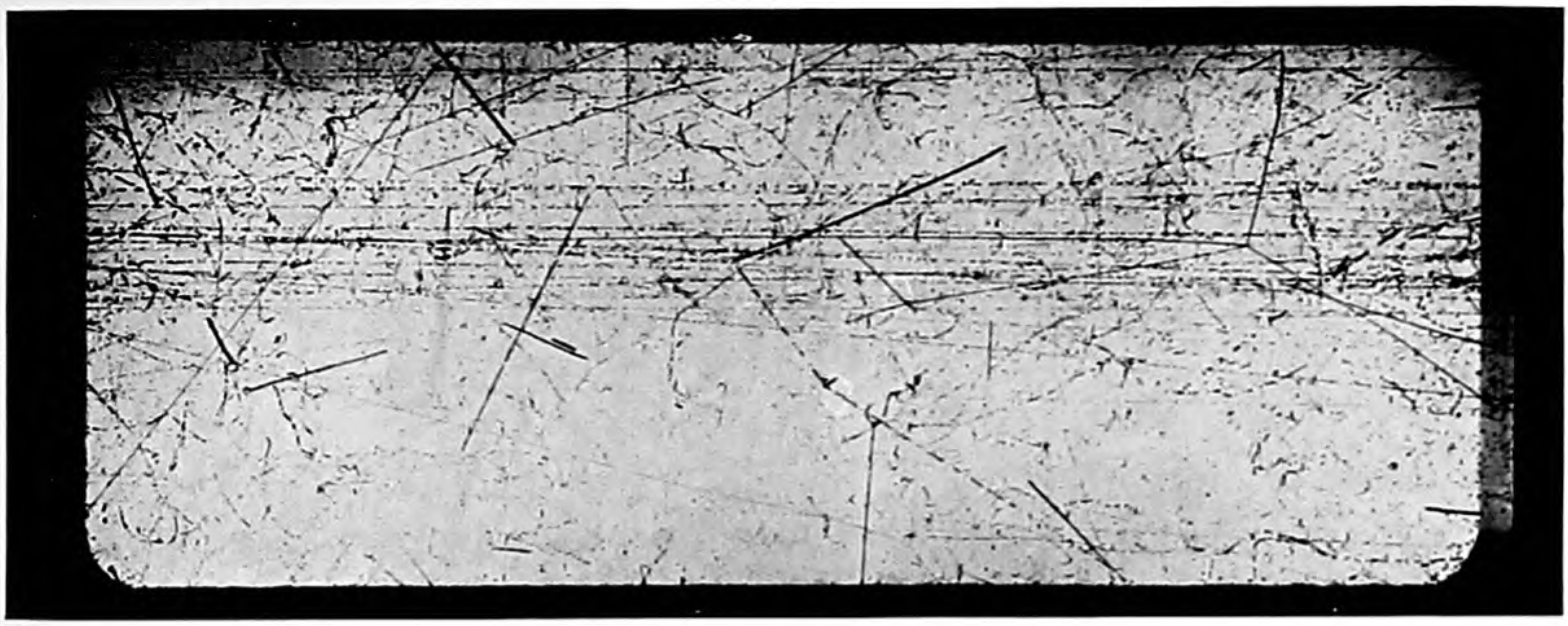
Conservation of parity was a law of quantum physics which said that there is no absolute distinction in nature between right and left. Experiments now show that such a distinction exists.

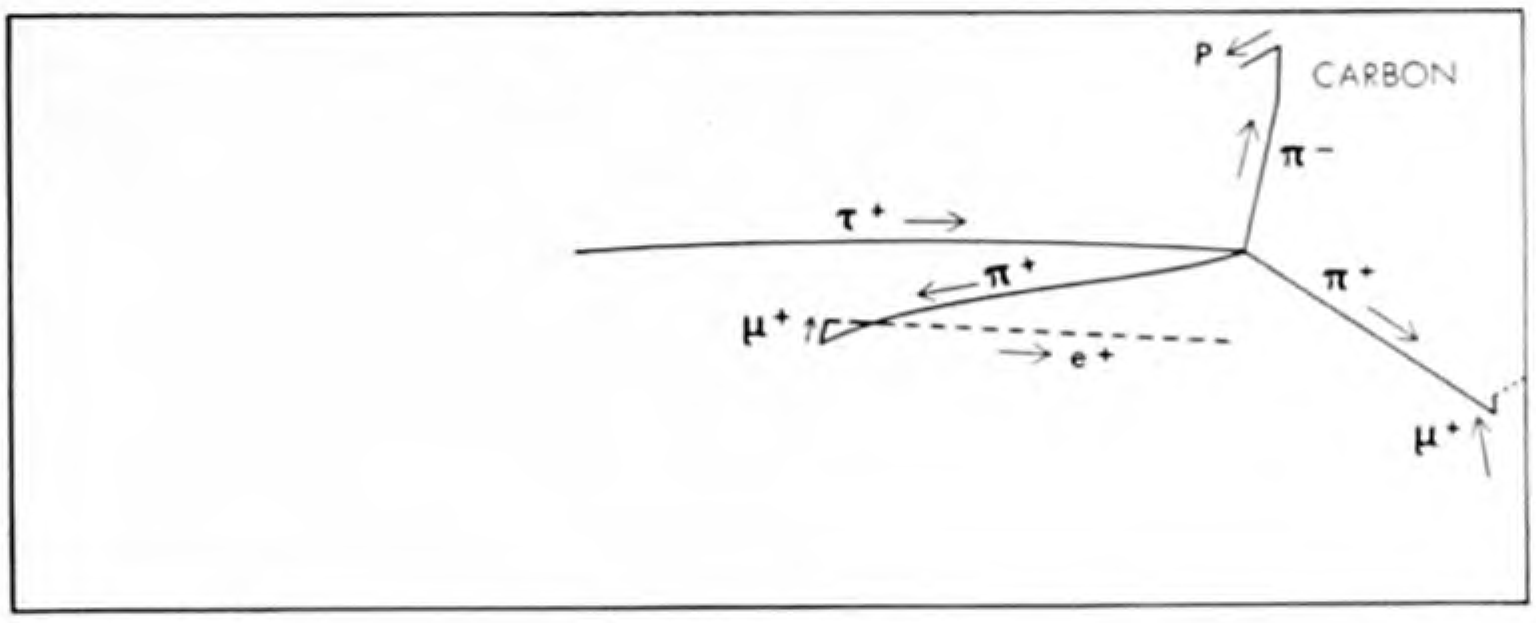
In the days when philosophers, acute in observation but as yet unaided by the tools of modern science, were primary founts of insight into the nature of the physical world, the philosopher Gottfried Wilhelm von Leibniz formulated a "great principle" which was to bear greater fruits than he knew. It was a proposition which at first thought seems absurdly simple and self-evident: namely, that "two states indiscernible from each other are the same state." Leibniz argued it on grounds which today we would find theological rather than scientific. Yet it has become one of the firmest pillars of modern physics. It underlies the theory of relativity and those laws of conservation—of energy, momentum and so on—upon which our understanding of nature is built. And it is now given deeper and sharper mean-



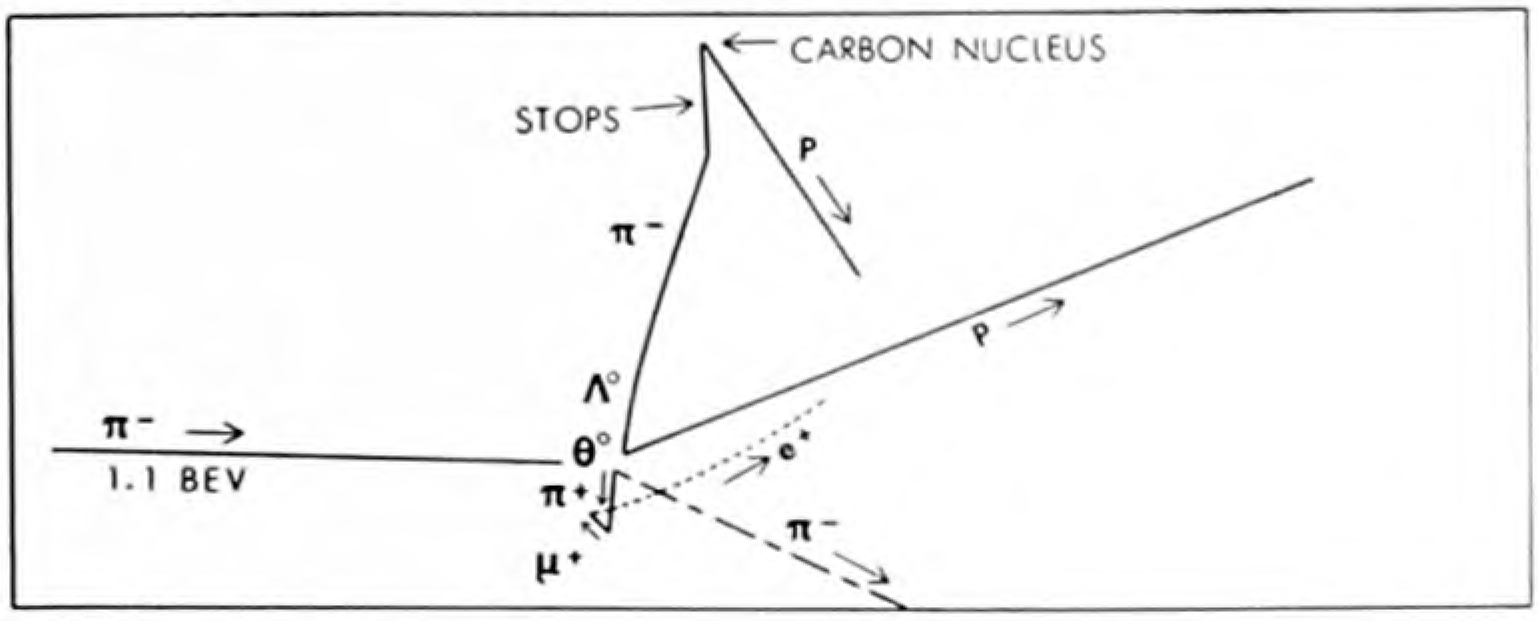


PI AND MU MESON DECAYS are processes in which parity is not conserved. This photograph records an example of both events. It shows bubble tracks made by particles from the Brookhaven Cosmotron in a chamber of liquid propane. As shown in the drawing at right, a pi meson (π^+) enters at left and decays into a mu meson (μ^+) about halfway across the chamber. The mu meson then decays into a positive electron (e^+).





TAU PARTICLE (τ^+) decays into three pi mesons. The one shown here enters at the left and travels almost to the right-hand end before bursting into a three-pronged path. The lower prongs represent positive pions, the upper prong a negative one. The experiments shown were performed by Donald A. Glaser, J. L. Brown, D. I. Meyer and M. L. Perl of the University of Michigan.

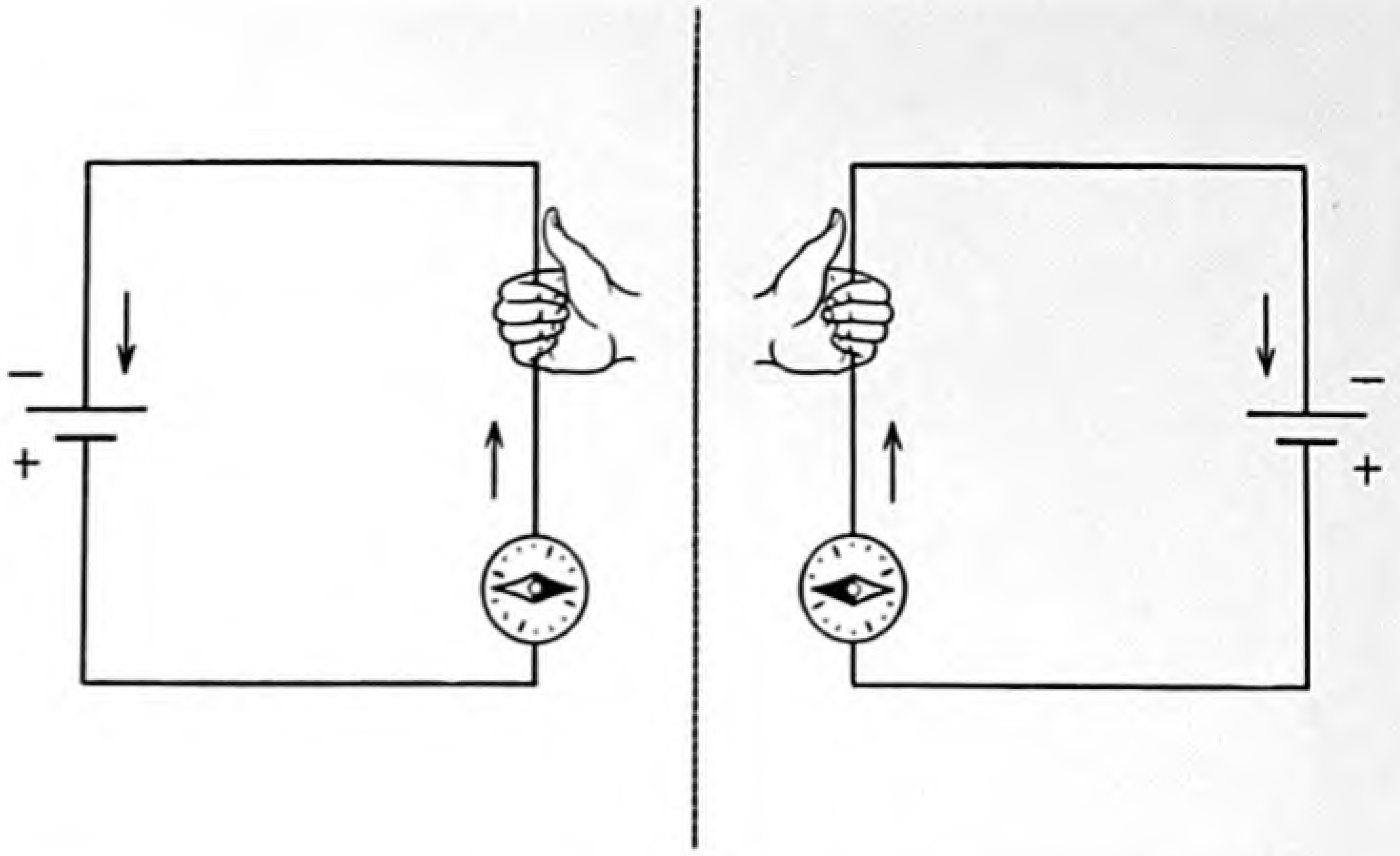


THETA PARTICLE (θ°) decays into two pi mesons. The theta in this photograph, being neutral, forms no bubbles and cannot be seen until it decays into a double prong. The theta decay is the lower of the two double-prong events shown in the drawing. This photograph also shows the birth of the theta particle. It was formed from the collision of a pi meson with a proton in the chamber.

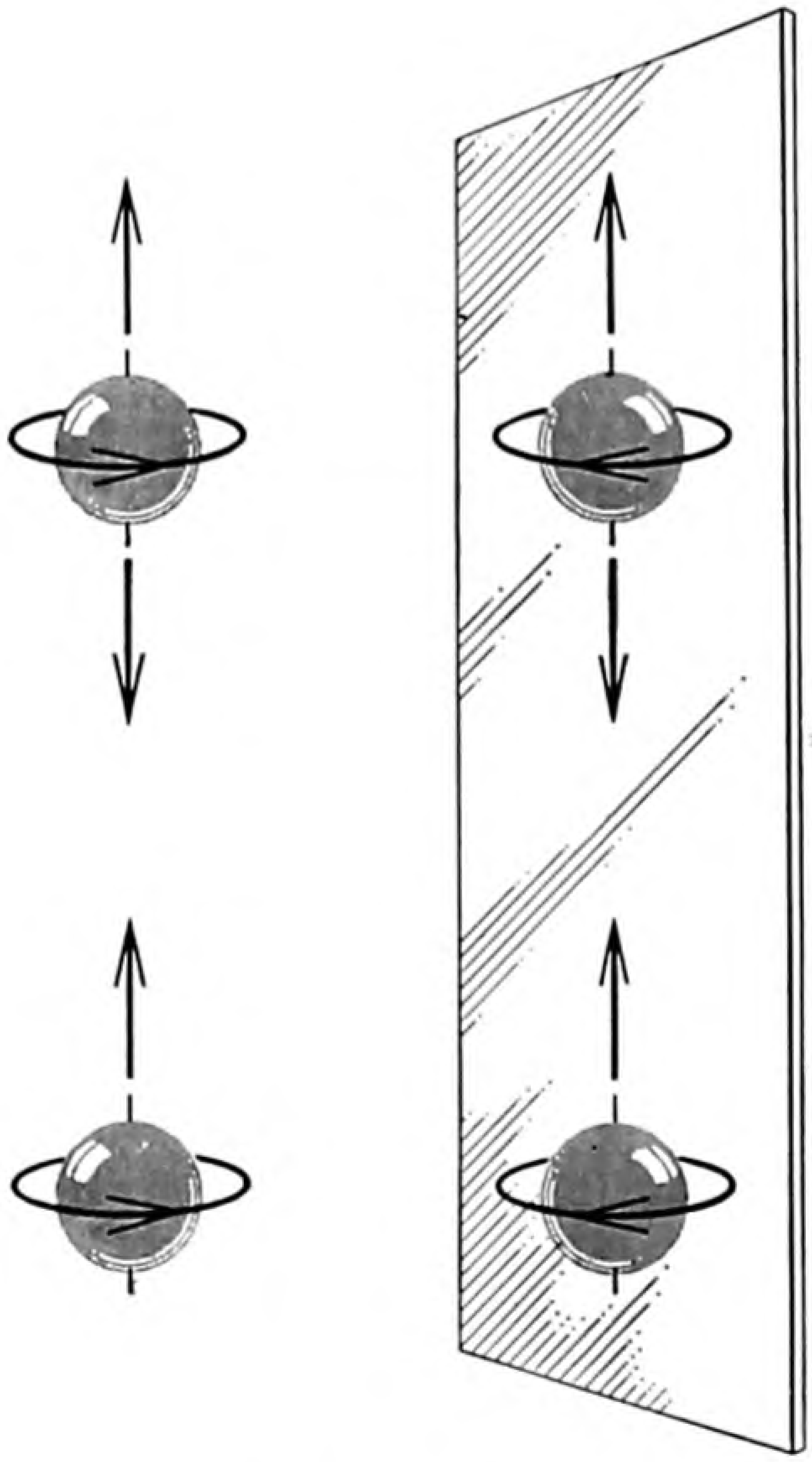
ing than ever by the amazing event in physics which is the topic of this article —the overthrow of the "parity principle" and the unraveling of the nature of left and right.

The important word in Leibniz's axiom is "indiscernible." Modern physics has been profoundly concerned with what is discernible and what is not. One of its strongest and most fruitful assumptions has been that among the indiscernibles are absolute space, time and direction. It is not hard to present examples. Think of the conventional world map. To each place are assigned a latitude and a longitude—a pair of numbers. The numbers are of great utility and convenience, but they are in no sense real attributes of the places; they have no physical significance. If the starting point for counting were to be shifted from Greenwich to Timbuktu, the numbers would change but no mountains would be moved. The numbers are merely arbitrary labels. And this is the manner in which space in general is treated in physics. The coordinates specifying positions in space describe only relative positions. We try to formulate our physical laws by the use of mathematical schemes in which absolute positions in space never enter. Whatever our frame of reference, we say, space remains invariant.

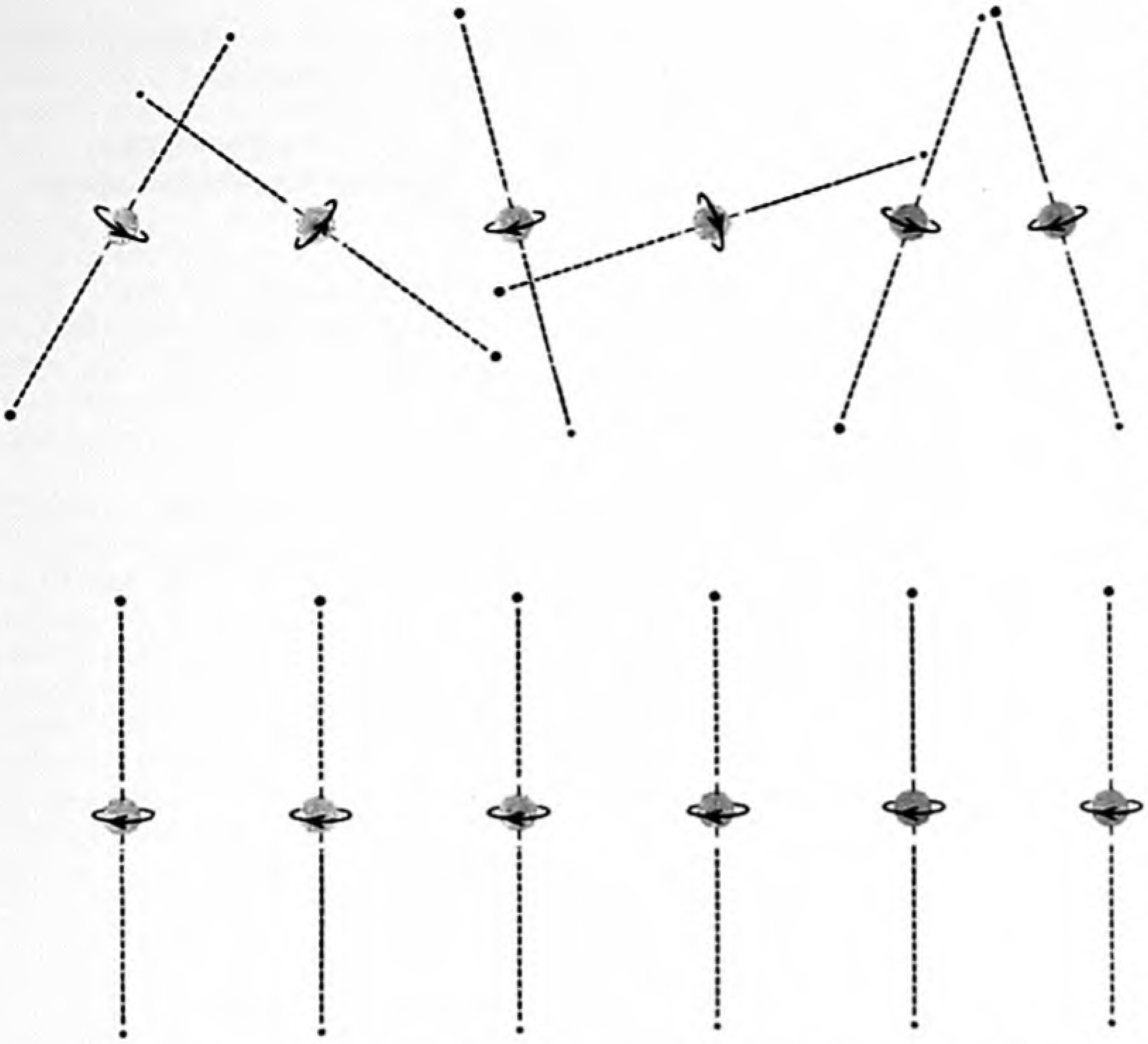
Let us take a more dramatic and comprehensive example. Suppose that a skilled director is going to produce on a stage before you some physical event or phenomenon—any whatever—without offering you clues to the date of the event, the directional orientation of the stage or the location of the theater. Could you determine any of these by any certain evidence? Indeed not. You may, of course, date the performance as lying within your lifetime, but this is clearly a subjective (*i.e.*, relative) time. (Indeed, Rip van Winkle could not succeed even in that.) You can judge which direction within the theater is up and which down, but "up" and "down" are merely relative to the earth; consider that you are in a theater in Australia and you will begin to realize the problem of attempting to determine the absolute orientation. A sharper test would be to use a compass to find "north," but this fails too, for the director can falsify the magnetic field, and in any case locating "north" on the earth tells you nothing about your absolute orientation in space. Nor can you locate the theater, even if you can look out a window and see a familiar landmark or a familiar star. The earth itself moves, and so do all the ex-



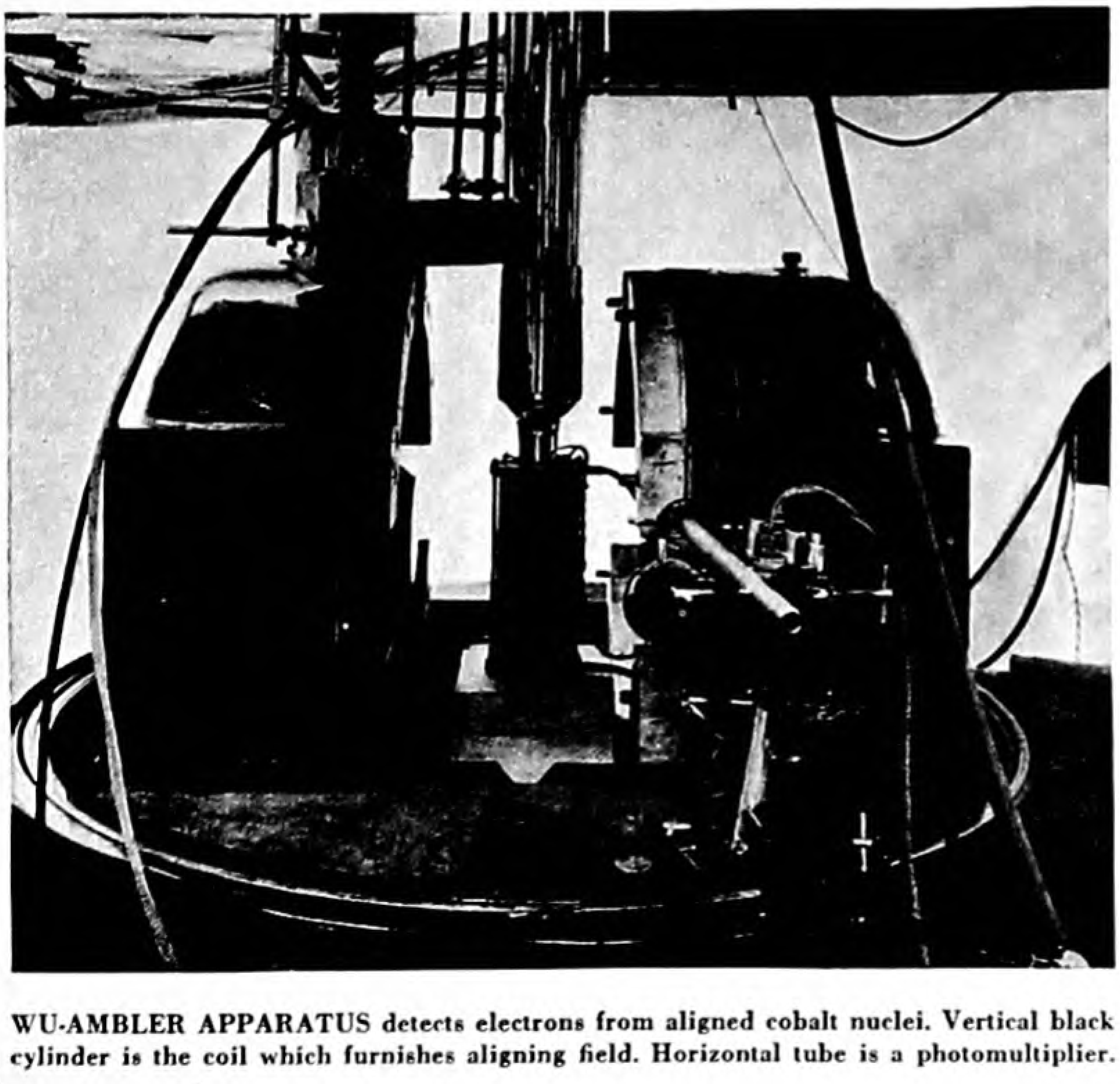
RIGHT-HAND RULE relating the directions of an electric current and its resulting magnetic field (thumb points with current, fingers point with the north-seeking pole of a test compass needle) becomes a left-hand rule when the experiment is reflected in a mirror. In this drawing, the "real" experiment is seen at the left and its mirror image at the right.



MIRROR REFLECTION of a spinning ball could not be detected if the ball ejected particles equally in both directions along its axis. Image at top right looks just like the real ball turned upside down. Reflection can be detected if there is a preferred direction for the ejection of particles. Thus the image at bottom cannot be mistaken for the real thing.



SPINNING NUCLEI of cobalt 60 (*gray circles*) emit electrons (*large dots*) preferentially in one direction along their spin axes and anti-neutrinos (*small dots*) in the other. The effect is ordinarily undetectable (*top*) but can be seen when spin axes are aligned (*bottom*).



WU-AMBLER APPARATUS detects electrons from aligned cobalt nuclei. Vertical black cylinder is the coil which furnishes aligning field. Horizontal tube is a photomultiplier.

ternal reference points. No, in principle and in practice absolute location in time or space and absolute direction are all indiscernibles.

These facts are of basic importance in physics. The indiscernibility of absolute coordinates lies at the basis of Albert Einstein's construction of the special theory of relativity, as is emphasized by the very name of his theory. And the fact that physical equations cannot refer to absolute time, space or orientation leads logically, by mathematical reasoning which we need not review here, to the classical laws of the conservation of energy and momentum.

The conservation "law" that concerns us now is the conservation of "parity," which rests upon the assumed indiscernibility of right from left. The indiscernibility principle can be put this way: there is no absolute distinction between a real object (or event) and its mirror image. A right-hand glove and a left-hand glove are surely different, but if accurately made they are precise counterparts; looking at a right-hand glove in a mirror, you cannot tell from its properties that you are not looking at a left-hand glove. The looking-glass world of people is admittedly unusual. Your mirror image is badly brought up: it offers to greet you with its left hand instead of its right, and it writes oddly. But the curiousness of the mirror world is largely conventional. There is no reason to doubt that such a world could exist. Indeed, any good director could set and coach a performance so that you could not distinguish it from the mirror image of a conventional performance.

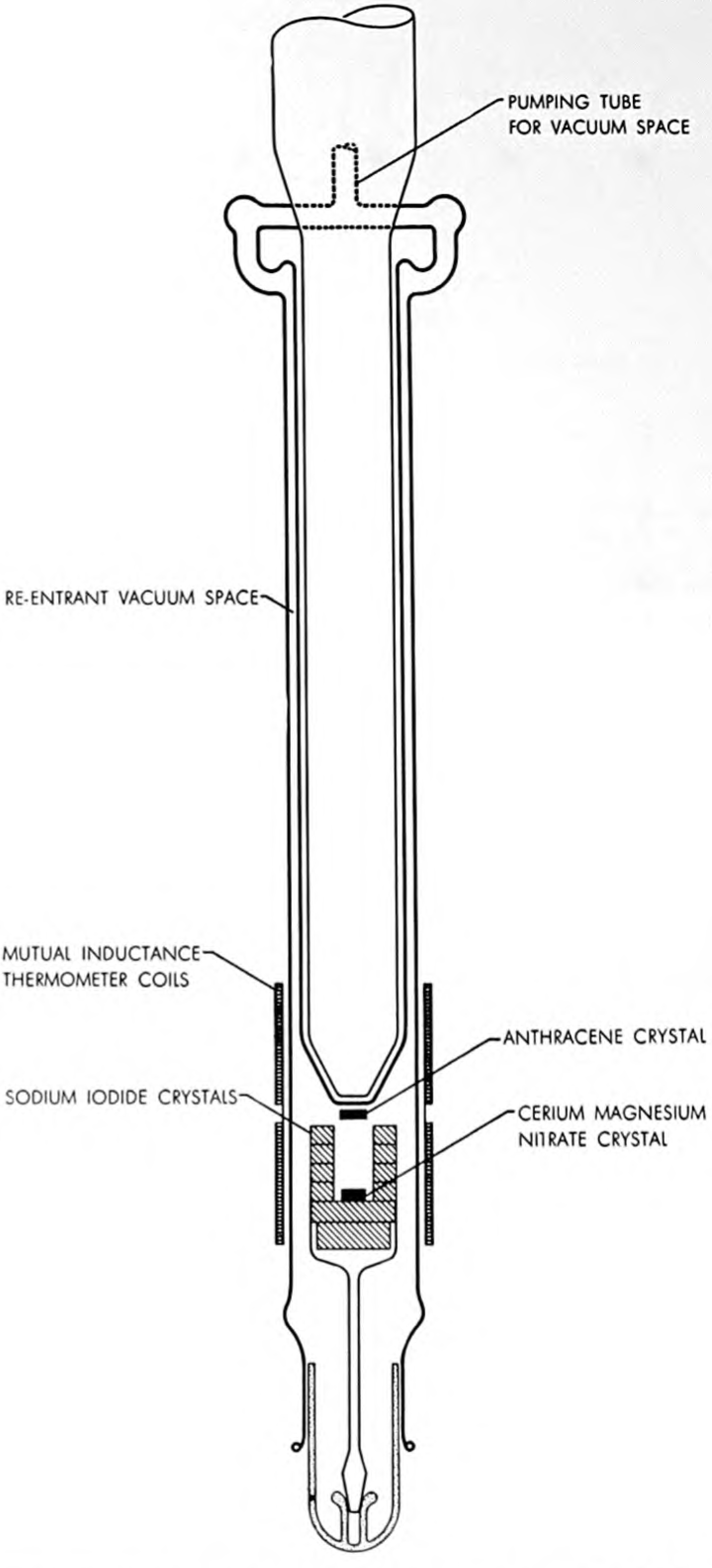
The physical principle of the indistinguishability of right and left implies that for every scene, every experiment, an exact mirror counterpart is possible. It is true that nature seems to favor a specific orientation (right-handed or left-handed) for the spiraling shells of snails and other animals, and for molecules of living matter. But this provides no conclusive evidence for the physicist: he sees that the molecule would function exactly the same way were it mirrored, and he can envision a world of living beings just like our own in which the "handedness" was simply reversed. The looking-glass world of life could function exactly like the actual world.

Until the startling events of the past few months, it seemed that the invariance of left and right was as unassailable as the invariance of time or space. All experience had buttressed the idea that no intrinsic difference could be found to

distinguish physical phenomena in a looking-glass world from the real world. All the seeming guides for distinguishing left from right failed on analysis to do any such thing. A student of electricity might cite the famous right-hand rule for telling the direction of an electromagnetic field: if you grasp a wire with your right hand so that the thumb points in the direction of the current's flow, your finger tips point in the direction in which a north pole moves in the magnetic field around the wire. But not so fast. What distinguishes a north pole from a south pole? True, you may refer me to a compass needle. The "north"-seeking end of the needle is colored blue and stamped with a big N. But this convention could be reversed without the least difficulty. If we switched the label N to the other end of the needle, the experiment would work with the left hand. So only convention distinguishes our experiment from its mirror counterpart. No intrinsic physical distinction, either in the macroscopic or the microscopic world, exists between a north magnetic pole and a south. All our right-hand rules are pure convention: there is nothing in the laws of electromagnetic fields that permits an absolute distinction between right and left.

The power of this idea of spatial symmetry ought now to be clear. If we begin by postulating mirror invariance, we can infer the indistinguishability of magnetic north and south poles. The latter in turn is an important principle in the microworld of particles, permitting us to give an orderly account of certain phenomena in that world. It implies the principle of the conservation of "parity," just as the invariance of space and time implies the conservation of energy.

"Parity" is a mathematical concept, impossible to define in physical terms. It is a property of the so-called wave function by which quantum mechanics describes the wave characteristics of a particle and represents its position in space. The variables of a wave function are just those coordinates that we use to locate spatial positions. Now it is not hard to see that if we change the sign of one coordinate (*i.e.*, from plus to minus), this is equivalent to reflecting the system in a mirror. Parity is the term that describes the effect of such a reversal upon the wave function. If the wave function remains unchanged when the sign of one of its three spatial variables is reversed, we say that the function has "even" parity. If reversal of the sign of the variable reverses the sign of the wave function, we call its parity "odd." In short, parity has one of two values—even or odd. And

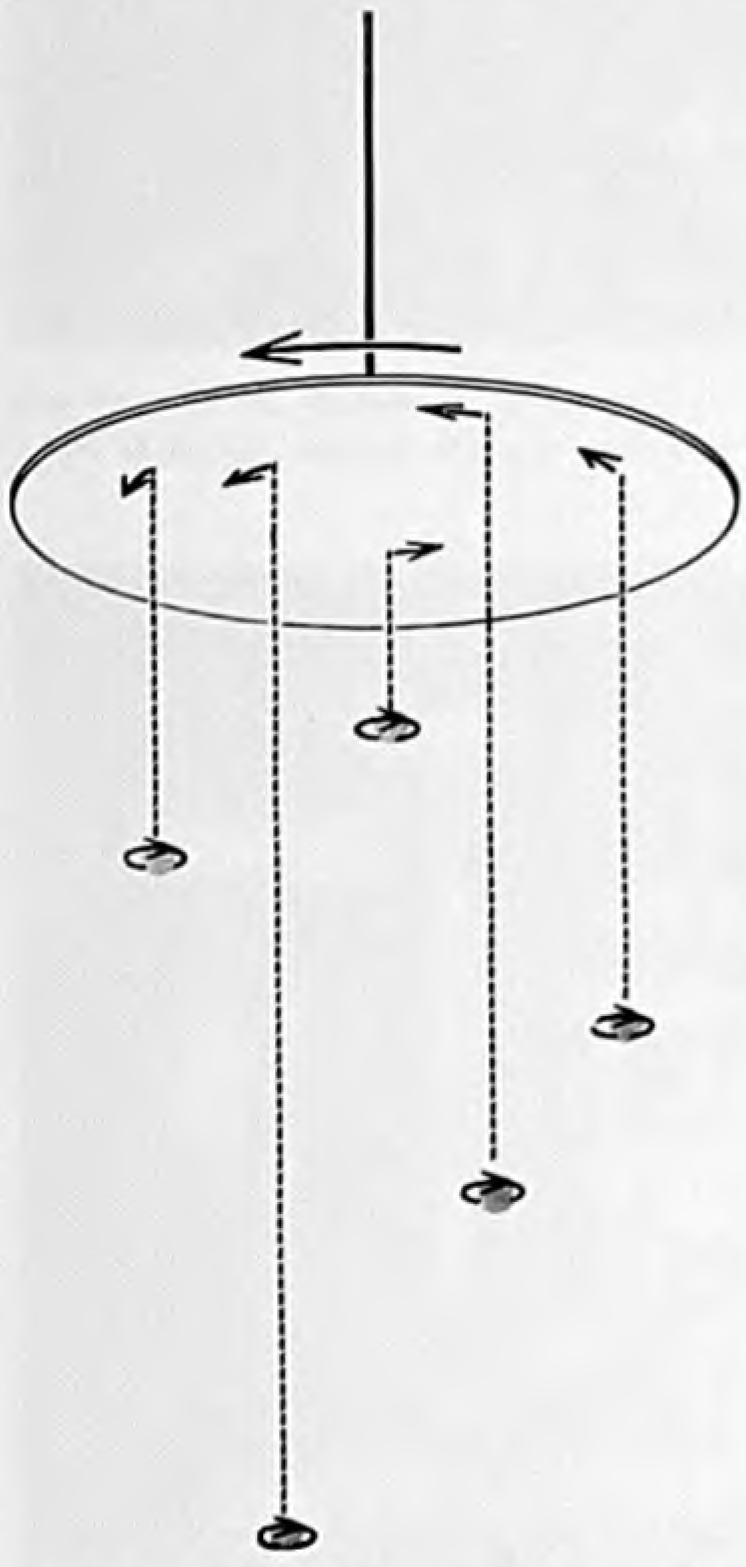


VACUUM CHAMBER for cobalt experiment is diagrammed above. The cobalt layer is deposited on the top of the cerium magnesium nitrate crystal. Anthracene crystal flashes when electrons strike it. Sodium iodide crystals count gamma rays from the cobalt, and indicate the degree of alignment. The large finger above the anthracene is a lucite rod which conducts the light of anthracene scintillations to amplifying and counting system.

all our experience, as well as theory, has indicated that in an isolated system parity never changes its value—*i.e.*, parity is always conserved.

Or rather, almost all. We are now confronted with a flat failure of parity conservation. The story goes back to about a year ago, when two extremely imaginative and ingenious investigators of the newly discovered "strange" particles of the atomic world made a really novel suggestion.

The two—Tsung Dao Lee of Columbia University and Chen Ning Yang of the Institute for Advanced Study—were absorbed in what was perhaps the most baffling paradox in this new realm of strangeness: the so-called "tau-theta puzzle." There were two mesons, called tau and theta. Tau, in the course of time, disintegrated into three pi mesons; theta, into two pi mesons. What was baffling was that in every property except the mode of decay, tau and theta were identical twins. Could they be one and the same particle? Decay of a particle by two different modes was certainly permitted by theory and precedent, but in this case the principle of conservation of parity stood in the way. Tau decayed to a set of pions of odd parity, theta to even-parity pions. The law of unchanging parity said that tau and theta must have different parity and therefore be different particles.



SCREW DIRECTION could theoretically be specified by an aluminum disk, mounted to rotate around vertical axis, and coated on underside with cobalt 60. Electrons ejected downward have predominantly one spin, and impart opposite spin to disk. Upward electrons are absorbed by aluminum. Thus the disk always spins in direction shown. The device was suggested by J. R. Zacharias of the Massachusetts Institute of Technology.

Yet the undownable question remained: Why were tau and theta exactly alike in every respect except this one? Lee and Yang boldly faced up to an embarrassing but insistent possibility: perhaps the parity-conservation law simply broke down in the realm of particle decays like tau's and theta's!

Their boldness was not rash. They could take the stand that while mirror invariance, from which the parity-conservation idea was derived, might hold in all other realms, it need not necessarily apply in the world of tau and theta. For the decay of tau and theta belongs to a very special class of reaction known as "weak interactions." The forces involved in them are very weak indeed—much weaker even than the forces which bind electrons in atoms. The forces are measured by the time it takes, with a given amount of available energy, for a particle emission to occur. By this measure the force entailed in a weak interaction such as the decay of tau or theta is smaller by a factor of 100 billion than the binding force on an atomic electron. Yang and Lee felt that the previous tests of mirror invariance in other fields of phenomena might have no validity in this untested realm of weak and subtle interactions.

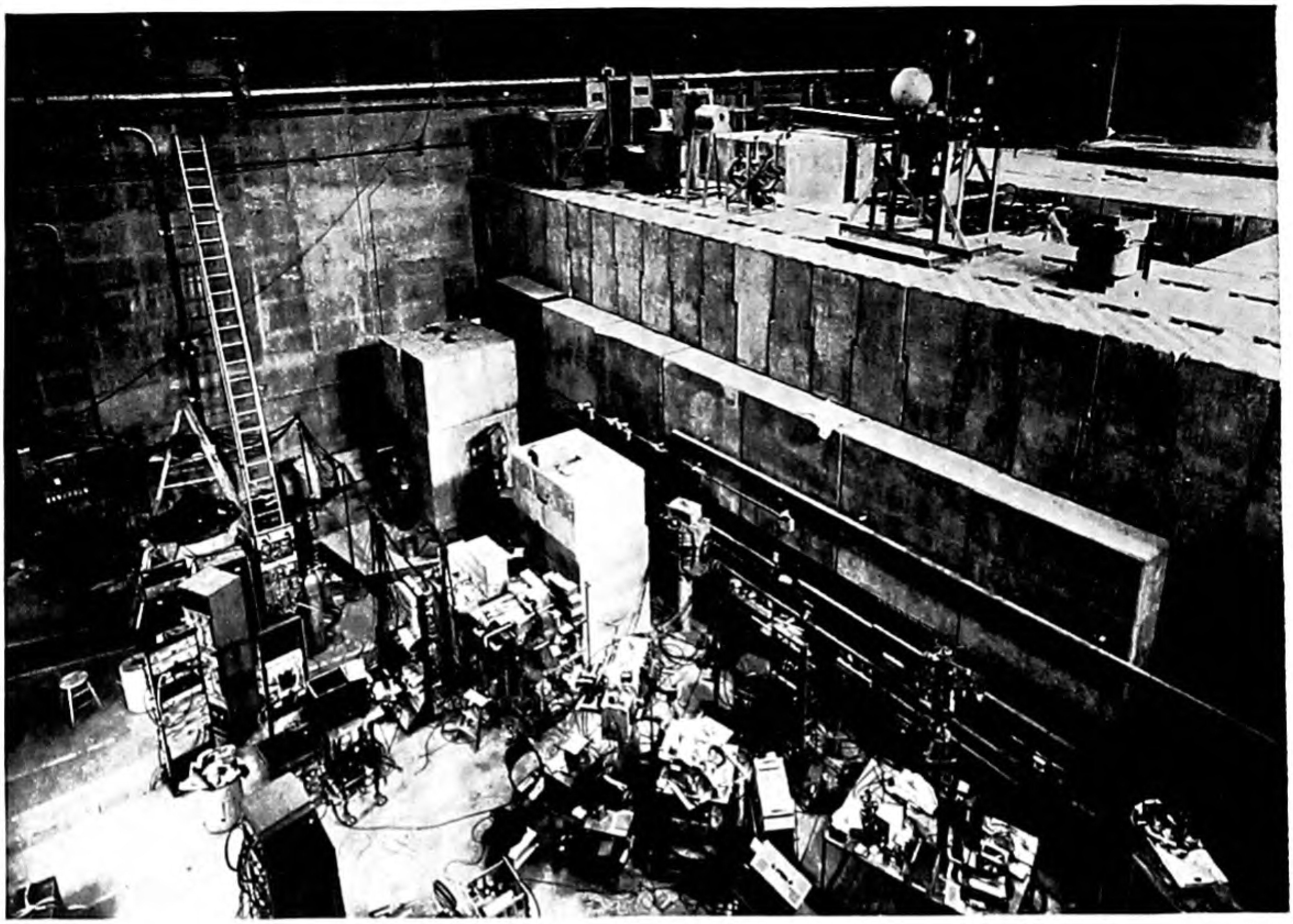
They proposed an experiment to test whether right and left could or could not be distinguished in this realm. Tau and theta particles themselves were poor candidates as subjects for such a test, for their lifetime is short—only about a billionth of a second. But the beta-decay (emission of beta particles) of radioactive atomic nuclei also belongs to the family of weak interactions, and these decays, taking place at much lower levels of available energy, have conveniently long lifetimes—measured in seconds or even years (*e.g.*, the beta-decay half life of cobalt 60 is 5.3 years). In essence Yang and Lee's proposal for the test experiment was simply to line up the spins of beta-emitting nuclei along the same axis, and then see whether the beta particles were emitted preferentially in one direction or the other along the axis. Neither direction would have any significance, conventional or otherwise: there is no arrowhead on the axis. But the preferred direction of beta-emission by the spinning nuclei would in fact define an arrowhead—the direction of advance of a right-hand or left-hand screw—and its mirror image would be discernibly reversed from the real thing. The experiment should leave no doubt about a distinction between left and right in beta-decay.

A powerful team of laboratory experimenters took up the challenge. Chien Shiung Wu of Columbia contributed her art in designing experiments and her experience in beta-decay work. A team at the National Bureau of Standards under Ernest Ambler undertook the task of lining up the nuclei. Ambler's job was to provide the straight line; Wu's, to look for the tip of the arrow (*i.e.*, the preferred direction of beta-emission).

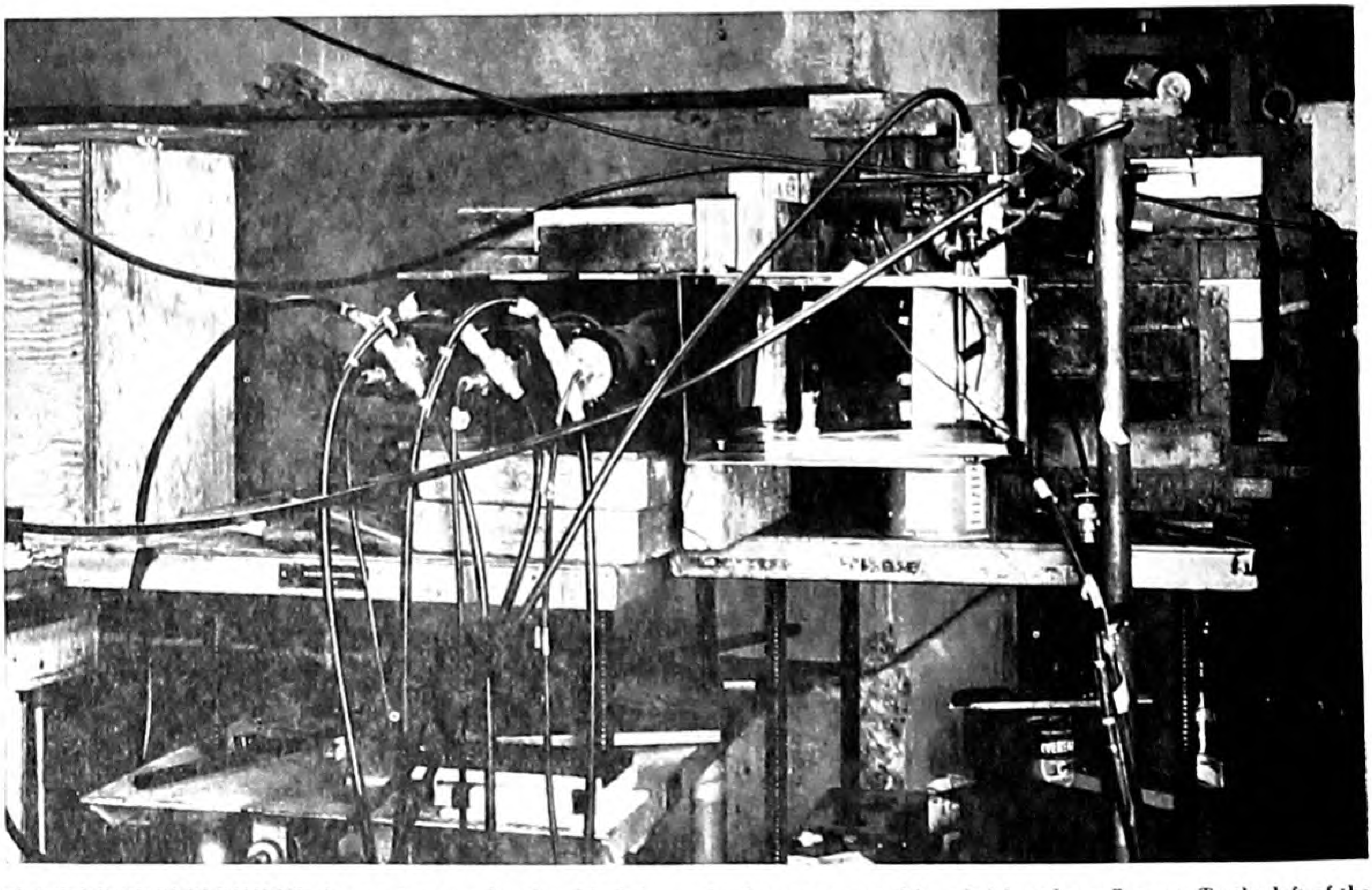
Nuclear alignment is a new art, no more than three or four years old. The sole handle by which the nuclei of atoms can be manipulated is their magnetic moment. No laboratory generator can produce magnetic fields strong enough to align these tiny moments; only within atoms themselves do sufficiently strong fields exist. So special atoms are lined up to produce a field, and their field in turn lines up the magnetic nuclei. But to make orderly alignment possible at all, thermal agitation of the atoms must be reduced to a minimum, which means cooling the system to very low temperature—considerably less than one degree above absolute zero. Shielded by the best sort of vacuum bottle, cooled by streams of liquid helium, the cobalt 60 atoms which served as the beta-emitters were kept aligned for 15 minutes or so at a time.

Six months to design, prepare and carry out the 15-minute experiments—these times proved just long enough to settle the issue beyond doubt. The beta particles emitted by the lined-up cobalt nuclei went predominantly in the direction against the magnetic field. This meant that, from the standpoint of beta-emission, the nuclei had an intrinsically left-handed spin. Left could be distinguished from right. Mirror invariance was dead. However valid it was elsewhere, in the realm of weak interactions it unambiguously failed.

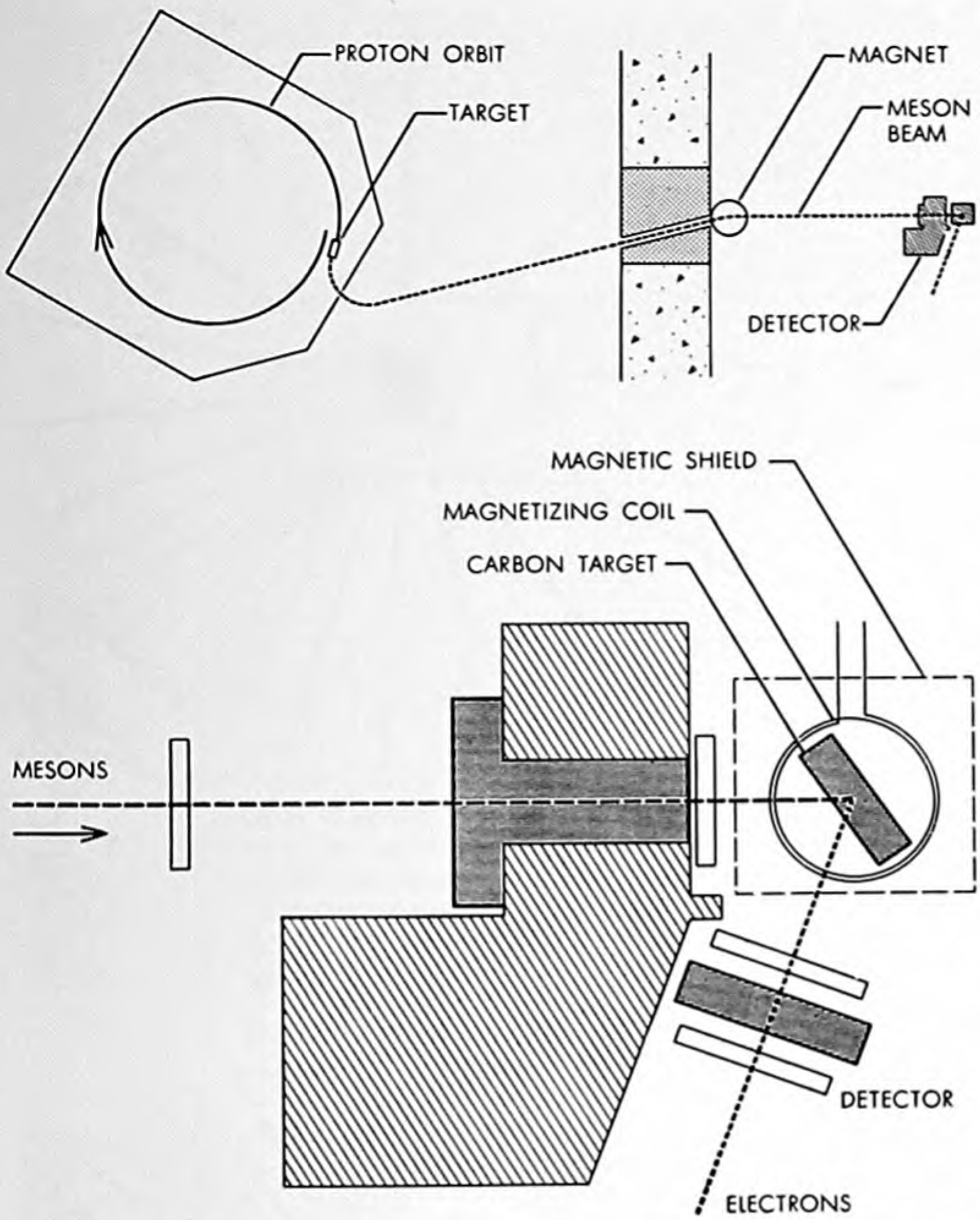
Within a few weeks after that first test in December, 1956, the conclusion was again unequivocally confirmed by another experiment. This time the weak



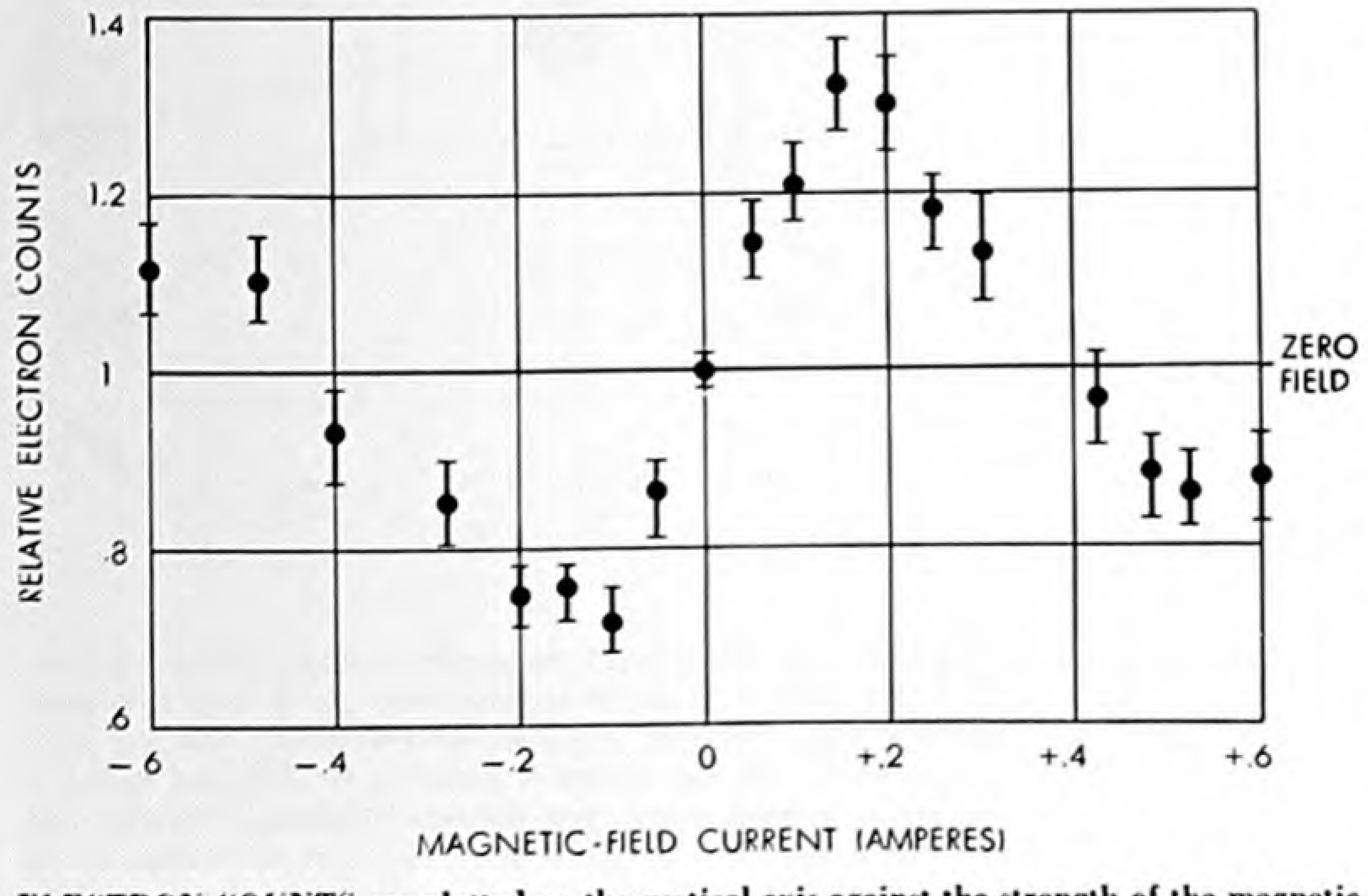
MESON DECAY EXPERIMENT, also showing lack of mirror symmetry, was done at the Columbia University cyclotron. Experimental setup is at center, between the small stepladder and the tall pile of concrete blocks. The cyclotron itself is behind shield at right.



CLOSE-UP OF EXPERIMENT shows wire-wound carbon block in which mu mesons are brought to rest. The block is in the rectangular frame supported by a brick and a coffee can. To the left of the frame is a triple counter which detects the emitted electrons.



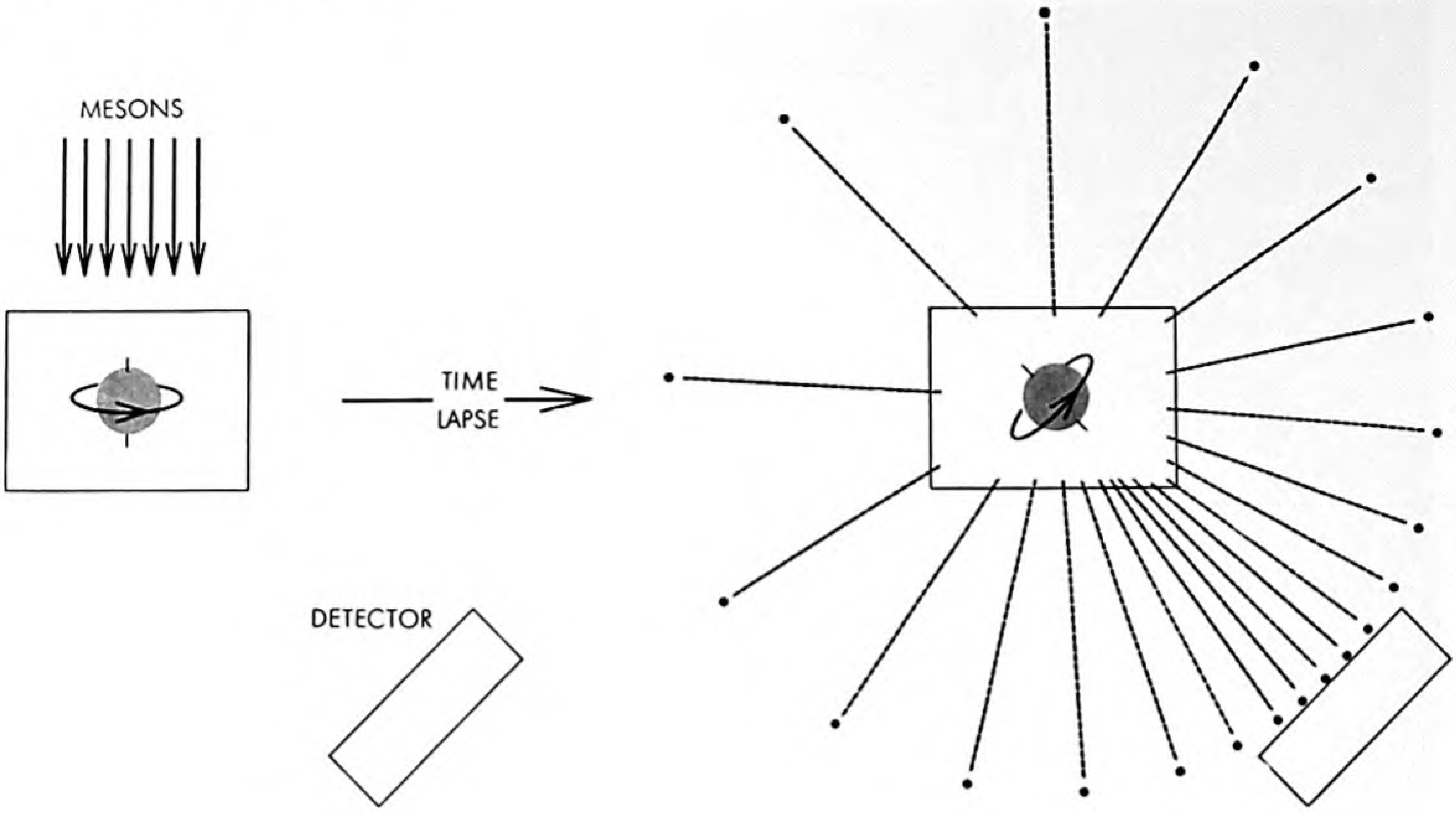
PATH OF MESONS, from accelerator to detector, is shown at top. At bottom is a detailed diagram of target and detector. T-shaped block stops pi mesons, leaving beam of mu mesons. These come to rest in carbon target and decay to electrons which are counted in the detector.

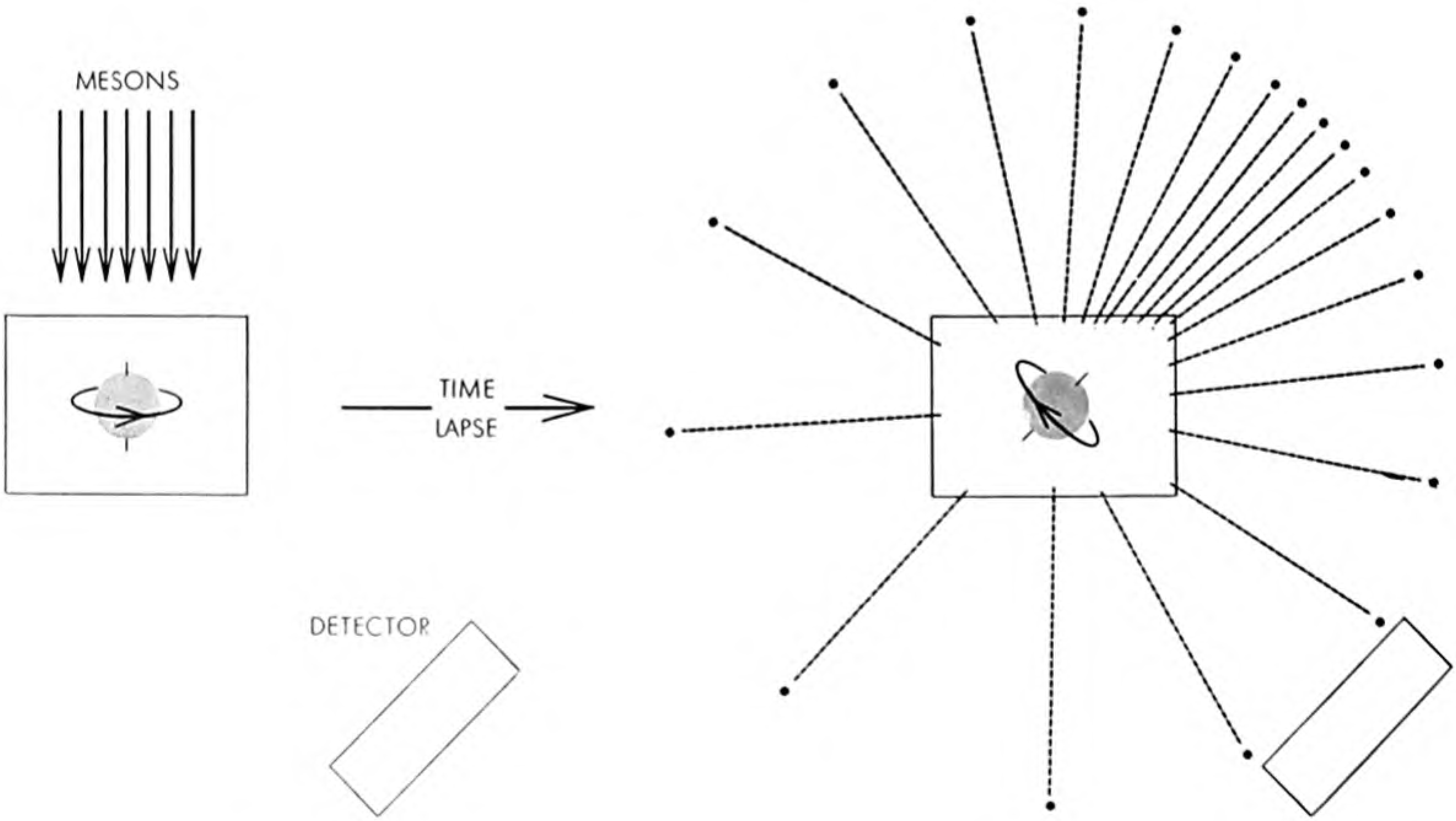


ELECTRON COUNTS are plotted on the vertical axis against the strength of the magnetic field which turns the electron-emitting mu mesons. Variation shows asymmetry of emission.

interaction tested was the decay of the mu meson. The experiment rested on a hypothesis which seeks to explain the failure of mirror invariance in beta-decay. The theory assumes that in the decay of a mu meson both a neutrino and an anti-neutrino are emitted, along with the beta particle, and that the neutrino always has a right-handed spin while the anti-neutrino's is left-handed. It was reasoned that when a pi meson decays into a mu meson and an anti-neutrino, the mu meson must be emitted with a left-handed spin to balance that of the anti-neutrino. As a consequence, when aligned mu mesons, under suitable conditions, decay with emission of electrons, the electrons should come out in a preferential direction. The theory was tested, first by a group at Columbia and later in other laboratories, and the preference for the specified screw direction was verified. The Columbia group have used the effect very ingeniously to measure the magnetic moment of the mu meson, and thus already have made a useful application of the failure of mirror invariance.

Theoretical physicists have only begun to speculate about the more general implications of this profound overthrow of a basic principle in the world of weak interactions. But there is an over-all lesson which can be put simply: The great invariance principles of nature may be relied upon within the domains of their application, but they are not *a priori* self-evident or necessarily of universal application. It is worth while to test to higher and higher precision the great foundation principles, including the conservation of energy. So far as we have gone, even in weak particle interactions, energy conservation appears to hold, but does it still hold for the weakest interactions of all, those involving the weak force of gravity? Here one thinks of the hypothesis that matter may arise spontaneously from a space containing no energy, and the possibilities are exciting. It may also be that there is some connection between the two major asymmetries we now see in the physical world—the right-left asymmetry of weak particle reactions and the fact that our world is overwhelmingly made up of one kind of matter, to the near-exclusion of anti-matter. Perhaps this lead could forge a bridge between the microphysics of the fundamental particles and the physics of the great distances—that is, cosmology. It is fair to say that the discovery of the limitations of the mirror invariance principle is not a setback but an opportunity. We have entered an exhilarating time.





ASYMMETRICAL MESON DECAY detected in the Columbia experiment is illustrated schematically. Pi mesons from the cyclotron decay into mus, and the mu mesons turn out to have their spins lined up with respect to their flight direction. They are stopped in a carbon block. Their alignment in the block is represented symbolically in the diagram by a single rotating sphere. A coil around the block produces a magnetic field which turns the spinning mu mesons. The field is allowed to act for about as long a time as it takes the average mu meson to decay into an electron. The decay electrons, shown as small dots, go off in all directions, but only those traveling in a particular direction from the block are counted. The experiment is repeated for different values of the magnetic field, which gave different amounts of turning. The electron counts vary as shown in the curve at the bottom of the preceding page. This shows more electrons are emitted along one direction of the mesons' spin axis than in any other direction.

The Author

PHILIP MORRISON is associate professor of physics at Cornell University, where he has done pioneer work applying physical theory in a number of fields, including microbiology. He graduated from the Carnegie Institute of Technology in 1936, then studied theoretical physics under J. Robert Oppenheimer at the University of California, where he received his doctorate in 1940. When World War II broke out, Morrison left a lectureship at the University of Illinois to join the Metallurgical Laboratory of the University of Chicago, and later became a group leader at the Los Alamos Laboratory of the Manhattan District. Morrison joined the Cornell faculty in 1946. His article "The Neutrino" appeared in the January, 1956, issue of this magazine.

Bibliography

CROSS PRODUCT AND CURL. Peter Gabriel Bergmann in *Basic Theories of Physics: Mechanics and Electrodynamics.* Prentice-Hall, Inc., 1949.

QUESTION OF PARITY CONSERVATION IN WEAK INTERACTIONS. T. D. Lee and C. N. Yang in *The Physical Review,* Vol. 104, No. 1, pages 254-258; October 1, 1956.

SCIENTIFIC AMERICAN October 1951, Vol. 185, No. 4, pp. 44-53 OFFPRINT 232

THE NEUTRON

by Philip and Emily Morrison

The celebrated fundamental particle has no electrical charge and the mass of 1,836 electrons. Although it is used to propagate the nuclear chain reaction and to probe the nature of matter, in itself it is an enigma.

THE periodic table, that ingenious roll call of the elements, lists them in the order of their nuclear electrical charge. They run from the lightest, hydrogen, with a charge equal in size but opposite in kind to the negative charge carried by one electron, to the heavy californium, with a positive charge of 98. The amount of positive electrical charge within the nucleus of an atom determines the number of its electrons, and the number and arrangement of atomic electrons in turn determines the chemical behavior of the elements. Chemistry is fundamentally the study of the atom's shell—the electrons, their arrangement and their laws of motion. It is the electrons alone that determine color, solubility and reactivity, magnetism and electrical conductivity, crystal structure, hardness and strength, indeed all the wonderfully varied properties of matter. The periodic table formed a rational basis for the study of chemistry even before there were elements known to fill all the places in the table. Now every charge number between 1 and 98 is accounted for, and the list is essentially complete; there may only be a few more heavy elements at the end, too unstable to be discovered in nature.

While the 98 elements of the periodic table are chemical elements, there is another that might well be added at the head of the list. There it would be the one element which belongs entirely to the physicist. We call it neutron. In some ways the neutron is the most interesting "element" of all. It has an electrical charge of zero, with no positive charge and hence no atomic electrons. Lacking a cortege of electrons, the neutron has no ordinary chemical or large-scale physical behavior—no color, no chemical compounds, no crystal structure. True, the neutron does interact with light waves, though they are far from the visible region, and it combines with nearly all elements as well as with other neutrons, but these reactions are so different in quality from those of the chemistry of the other elements that the two are not comparable. The neutron is all core and no shell; its behavior is purely nuclear. In an ordinary chemical reaction an atom of one element links with another by gaining or losing electrons from its outer shell, while the nucleus remains unchanged. But when a neutron combines with another element, the electrons are at first largely unaffected, and only the nucleus is changed.

The Neutron and the Nucleus

The electrons moving around the nucleus of an atom form a barrier of negative electrical charge which keeps its distance from the attracting nucleus, and holds the atom separate from its neighbors. Thus, by discouraging the interpenetration of atoms, electrons preserve the identity of different substances. It is true that one cannot store water very long in a bright iron can, because the atoms of iron will, by a series of more or less complicated steps, unite with the atoms of oxygen from the water to form a new atomic arrangement, molecules of the compound rust. Eventually, as enough rust is formed, the can will disintegrate entirely. It is possible, however, to keep water indefinitely in a glass jar and have almost no reaction whatever between the water and the container.

On the other hand, it would be difficult to design any vessel that would hold a gas composed entirely of neutrons for even a few thousandths of a second, since the particles of such a gas would instantly leak into the walls of the container and combine with its atoms. No electric charge can repel the neutron, and it reacts with the nuclei of nearly all elements. When a neutron passes through matter, it cannot be deflected from its course by ordinary electrical forces, but only by the forces that are specifically nuclear. These forces extend a mere 10^{-13} centimeters, a ten thousand billionth of a centimeter, into the space around the nucleus, or some hundred thousandth of the average distance from the nucleus to its outermost electron. A neutron hardly notices the shell of electrons surrounding each atom, but moves freely on its way, often for several centimeters, before it collides directly with a nucleus. Then the neutron may bounce off or stick tight, depending to some degree on the speed with which it is moving. After several collisions the neutron sticks to some nucleus; the neutron gas that had leaked into the walls of our container would gradually disappear as its particles were absorbed.

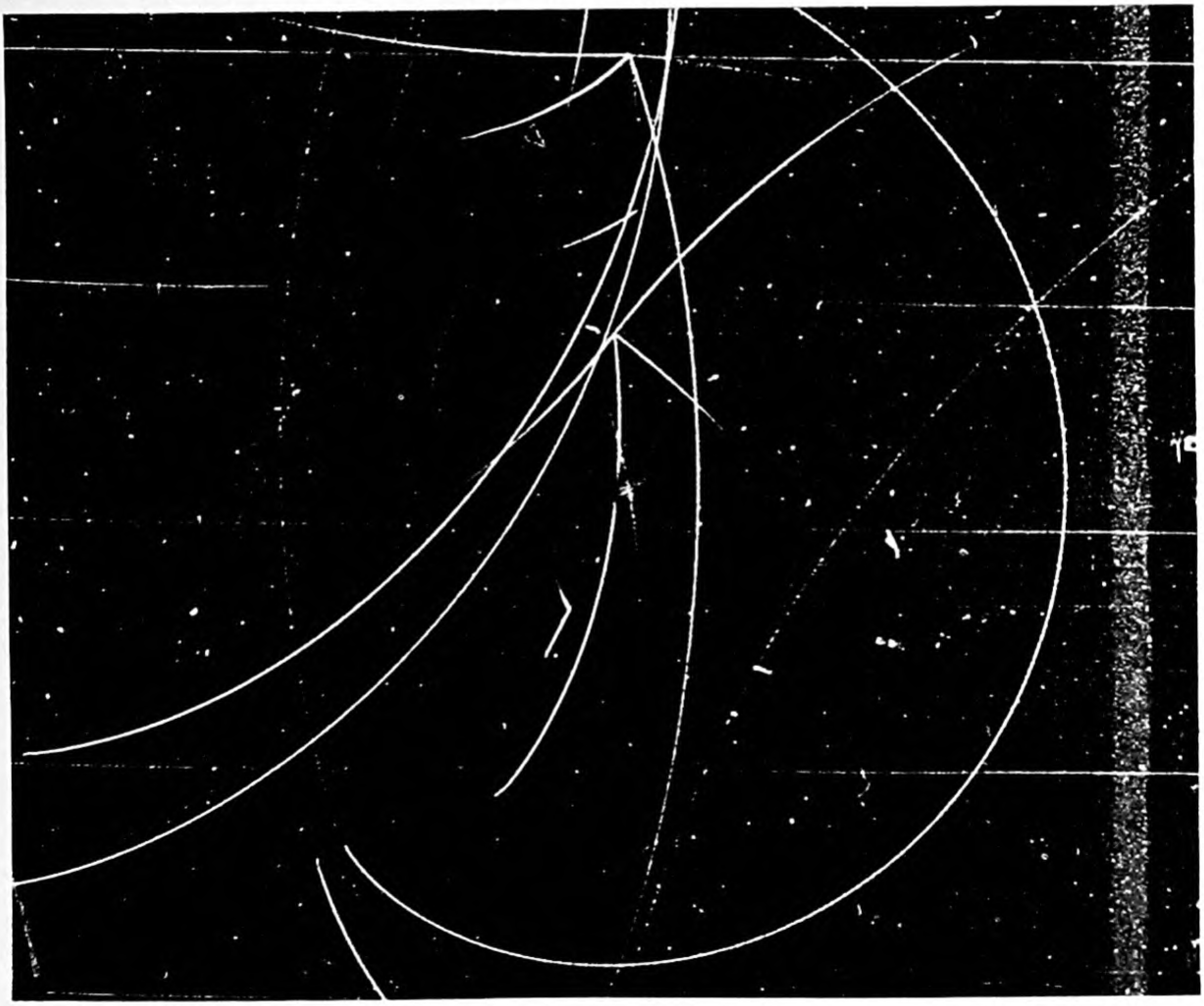
Just as any other gas, neutrons can at least in thought be compressed to form a liquid droplet of several particles touching each other, a material that would be as dense as anything we know. Since the diameter of the electron orbit of an atom is so much larger than the diameter of the nucleus, and since a neutron liquid has neither atomic electrons nor the volume they loosely occupy, a density of 10^{14}, or a hundred thousand billion, grams per cubic centimeter becomes plausible. A drop of neutrons the size of a drop of water would weigh more than the Washington Monument. The dense nucleus of every atom behaves like just such a tiny droplet of neutrons combined with protons. Ordinary solid matter thus seems to a neutron much like a good vacuum would seem to a moving charged particle. An electron can travel for several centimeters without hitting something only in a good vacuum in which the atoms are widely separated, but a mean free path of this length is commonplace for a neutron, even in solid material. For a charged particle, space is full of great whirling balloons—the atoms. For the neutron, the wall of each atom-balloon is porous and insubstantial, and the only solid obstacle is a tiny central grain of sand.

The Clue of the Isotopes

All the atoms of a given element are chemically identical, *i.e.*, they have the

same number of electrons. The weight of their nuclei, however, may differ. By separating from their fellows all atoms of the same weight we obtain a pure isotope. All elements have a number of such isotopes, stable or unstable. The neutron has no long-lived stable isotope, and probably no second, unstable isotope exists either, though a transient association of two neutrons may last for a very few cycles of its internal motion. The neutron itself is unstable—spontaneously radioactive in much the same way as some natural radioactive associates of uranium are. Instead of decaying like radium or uranium by emission of an alpha particle composed of two protons and two neutrons, the neutron decays into three particles: an electron, a proton and a neutrino. This process is called beta decay. It has been predicted theoretically that half of a given number of neutrons will decay in about 10 minutes, *i.e.*, the neutron has a half-life of 10 minutes. Since a neutron lasts only a few thousandths of a second before being absorbed into a nucleus of solid material, the number of neutrons that remain free long enough to decay spontaneously is extremely small and their detection most difficult. Recently some very elegant experiments have confirmed these predictions of theory. While the evidence for the beta decay theory is reasonably clear, the details are far from completely understood. An electron is certainly sent out as the neutron decays, but it is difficult to see how an electron can be inside a neutron in any understandable way, since the volume required by an electron is far larger than the entire volume of the neutron. The electron is created, with the enigmatic neutrino, on the instant of decay. Because most neutrons are absorbed quickly, and even those not absorbed into other nuclei would decay of their own accord in a short time, free neutrons are hard to find. Their natural habitat is nowhere. The rocks, the sea and the air are our sources for all the other elements, but there is no natural storehouse of neutrons. A few are manufactured high in the atmosphere as the result of collisions between cosmic rays and the nuclei of the atoms that compose the air, but only at the rate of about one neutron per square centimeter per minute. A few are made by reactions initiated by radioactivity in rocks. The first neutrons to be discovered were part of a steady stream actually produced in the laboratory.

The early study of isotopes had suggested several reasons why atoms of the same element varied in weight. Among the most popular of these explanations was the notion that the additional weight consisted of extra protons whose charge



CLOUD CHAMBER PHOTOGRAPH shows the effects of neutrons but not the neutrons themselves. The cloud chamber stands beside a cyclotron at the University of California. When 100-million-volt neutrons are directed into the chamber, some of them strike the nuclei of oxygen atoms and cause them to fly apart. The tracks of two such exploding nuclei are at top and center. Their fragments are curved by a magnetic field about the chamber.

was without effect because each was neutralized by an ordinary electron packed within the nucleus. This account was for many reasons unsatisfactory, and in 1920 the great British physicist Lord Rutherford presented his theory of a neutral particle to explain why nuclei were too heavy for their charge, a theory which foreshadowed with remarkable clarity the experiments of 10 years later. Said Rutherford in the Bakerian Lecture before the Royal Society of London:

"[This assumption] involves the idea of the possible existence of an atom of mass 1 which has zero nucleus charge. Such an atomic structure seems by no means impossible. On present views, the neutral hydrogen atom is regarded as a nucleus of unit charge with an electron attached at a distance, and the spectrum of hydrogen is ascribed to the movements of this distant electron. Under some conditions, however, it may be possible for an electron to combine much more closely with the H nucleus, forming a kind of neutral doublet. Such an atom would have very novel properties. Its external field would be practically zero, except very close to the nucleus, and in consequence it should be able to move freely through matter. Its presence would probably be difficult to detect by the spectroscope, and it may be impossible to contain it in a sealed vessel. On the other hand, it should enter readily the structure of atoms, and may . . . unite with the nucleus. . . ."

Beginning in 1930 a series of experiments in Heidelberg by W. Bothe and H. Becker and later in Paris by Frédéric Joliot and his wife Irène Curie produced a kind of uncharged radiation that had penetrating power that was curiously high even for gamma rays, the very short electromagnetic waves. As a source for these new rays the experimenters used a disk of beryllium metal, against which they directed the alpha particles constantly emitted by a member of the radium family. James Chadwick of Cambridge University repeated the experiment with the same kind of source, but directed the suspected new radiation into a Wilson cloud chamber filled with nitrogen. He observed that an occasional nitrogen atom would dart across the chamber as though it were recoiling from an impact much too massive to be caused by a weightless if energetic gamma ray. The events in his cloud chamber looked as though two particles as heavy as nuclei had collided and bounced apart, though the track of only one was visible. From this evidence Chadwick was able to prove that the new rays were not electromagnetic radiation but particles. The particles had a mass very similar to that of the proton but no electrical charge. For this latter reason the particles could not knock electrons from the atoms in the cloud chamber, as they would have to do in order to leave a visible trail. These electrically neutral particles Chadwick called neutrons.

The discovery of the neutron confirmed Rutherford's old prediction and made clear the origin of the extra weight in the nucleus: the protons provide the positive charge of the nucleus, and the neutrons provide the excess weight. Neutrons exist in the nucleus of every element except the single-proton common isotope of hydrogen. In light elements they supply about half of the mass, and in heavy elements they supply much more than half. The nuclei of the two stable isotopes of carbon each contain six protons, which identify them as carbon. But carbon 12 has six neutrons and carbon 13 has seven. The isotopes of uranium each have 92 protons, much less than half the weight of the uranium nucleus. To make up the difference uranium 235 has 143 neutrons, or 51 more neutrons than protons, and uranium 238 has 146 neutrons.

The Production of Neutrons

To study the neutron, as to study any other element, one must first purify it; in this case purification means separating the neutron not from any mere chemical combination but from a tight nuclear bond. Such a physical separation of the condensed nuclear material is far more difficult than even the most obstinate chemical separation, as, for example, that of the compounds of fluorine. From the chemist's point of view the extremely resistant fluorocarbon plastics are tightly bound and stable compounds, but with sufficient effort he can break them down with heat or molten alkaline reagents. The separation of nuclear material, however, involves wholly different techniques. The simplest neutron source is essentially the one used by its discoverers: an inch-long sealed brass capsule in which alpha particles from a radium salt bombard beryllium metal powder. Most of the fast alpha particles come to rest after frittering away their energy against the atomic electrons, never encountering other nuclei. But on rare occasions an alpha particle will collide with a beryllium nucleus and expel a neutron. A good yield is one neutron for every 5,000 alpha particles emitted by the radium. Such a nuclear reaction creates a comparatively small trickle of neutrons. In actual numbers such a trickle may amount to a million per second, but at this rate it would take all of geological time to set free even a single gram of neutrons!

Any source of fast particles or gamma rays with enough energy directed against a target made of a light element will set free neutrons. The prewar physicist typically used fast deuterons, composed of a neutron and a proton, on beryllium in a cyclotron. The intensity of neutrons produced by the deuteron beam of even a modest cyclotron equals that of many kilograms of radium. But if neutrons stream from the largest cyclotron, a veritable Niagara of them bursts from the chain-reacting atomic piles developed during the war. The largest neutron-producing plant of all is that at Hanford, the Atomic Energy Commission's plutonium factory in the state of Washington. Hanford manufactures neutrons as a nuclear reagent, and a few grams of neutron gas are sufficient to make a kilogram of plutonium. About half the weight of every lump of dirt is neutron, since, as we have noted above, the nucleus of almost every atom contains as many neutrons as protons. But a billion-dollar plant can produce in one day an amount of the pure nuclear reagent to be measured only by the ounce. Expensive stuff! Once made, however, this intensely powerful reagent is the philosopher's stone in the large-scale alchemy of the plutonium project

The Detection of Neutrons

It is a simple but not a trivial matter to detect the presence of neutrons, even in small quantities. Any fast charged particle leaves a wake of charged atoms, or ions, which may easily be recorded on a photographic plate or detected by such electronic devices as the Geiger-Müller counter. Since neutrons have no charge and therefore cannot knock free any atomic electrons, they leave no such track to be recorded. It is possible in principle to detect the accumulation of neutrons by measuring the gain in weight of the material surrounding the neutron source as neutrons are captured. Such a method is impractical, however, because of the extremely small weights involved. Even if one used hydrogen, whose individual atoms' weight would double on the addition of a neutron, there would be so few hydrogen atoms affected that their presence could hardly be noticed among all the others. For special investigations with the most copious neutron sources—nuclear reactors—such measurements can be made with a sensitive mass spectrograph, but even this method is too delicate and complex for any routine detection. For ordinary purposes neutrons may be detected only indirectly through some specifically nuclear effect. One important and typical method employs boron 10. When it captures a not-too-fast neutron—which it can do with a probability nearly unexcelled among nuclei—this boron isotope flies apart into two fragments, one a fast-moving alpha particle and the other a nucleus of lithium 7. Both the alpha particles and the lithium nuclei leave a strong, easily detectable track of ions in their wake. The basis for nuclear reactors and atomic

bombs is of course the fission of uranium 235 when it captures a neutron; this phenomenon can also be used to detect neutrons. The fission method of detection, with boron 10 or uranium 235, is most useful for rather slow neutrons. The detection of fast neutrons is often managed differently. A neutron traveling at high speed is not usually captured by the first nucleus it meets; it collides only to recoil. In the collision the neutron slows down, while the nucleus that was hit goes faster than before the encounter; the neutron has shared its kinetic energy with the nucleus it has bumped. By mere impact the fast neutron jolts the target nucleus sharply off its course, usually without disturbing it internally. As it bounds away from the collision, the nucleus leaves many of its electrons behind. Now that the atom has become a charged ion, it leaves a wake of other ions which can be detected. Almost every neutron comes finally to the end of its free life captured by some nucleus; this very efficiency provides still another method of detection. A small foil, say of silver, is exposed to a stream of neutrons. Each silver nucleus that acquires an extra neutron becomes unstable, and within a short time emits a beta particle whose ionizing track may be detected in some standard counter.

Let us trace the career of a free neutron from the moment of its liberation. To do this we may use an ordinary neutron source, a capsule filled with beryllium and radium, placed in a bucket of water. Occasionally, if the circumstances are favorable, a neutron is released from a beryllium nucleus as it is hit by an alpha particle. A steady supply comes out of the capsule in all directions. As each neutron emerges from its nuclear bonds, it is traveling at about a twentieth the velocity of light, with an energy of some millions of volts. The neutron passes through hundreds of millions of atoms in the inch or so it traverses before colliding with a nucleus of hydrogen or oxygen in the water. The fast neutron usually bounces off the nucleus in this first collision, going off in a new direction at a slightly lower speed, while the nucleus is jolted into moving faster than before. With each collision the neutron loses a little more of its original store of kinetic energy, which is taken up by the nucleus that it hits. The nuclei in the water have had the effect of moderating the speed of the neutrons without capturing them, and for this reason water is called a moderator. Many elements with light nuclei are effective moderators. Compounds of hydrogen, oxygen, carbon and beryllium are all satisfactory, but water, graphite and paraffin are most frequently used because they are readily available in pure form. After several hundred nuclear collisions, the neutron has slowed down until its motion is comparable to that of the relatively slow-moving atoms in the water around it. The neutron is still moving at supersonic speed, but this is thousands of times slower than when it was freshly released. Such a neutron which has come to share the ordinary heat motion of the atoms of the moderator is called a thermal or slow neutron. Sometimes a neutron will escape by chance through the wall of the bucket into the air of the room, where, of course, the molecules are far apart compared with those of water. Here the neutron can travel a long distance, often several hundred yards, before colliding with another nucleus. Most of the neutrons in pure water, once they have slowed down, will be finally absorbed by nuclei of hydrogen to form the isotope deuterium, famous as the essential element of heavy water.

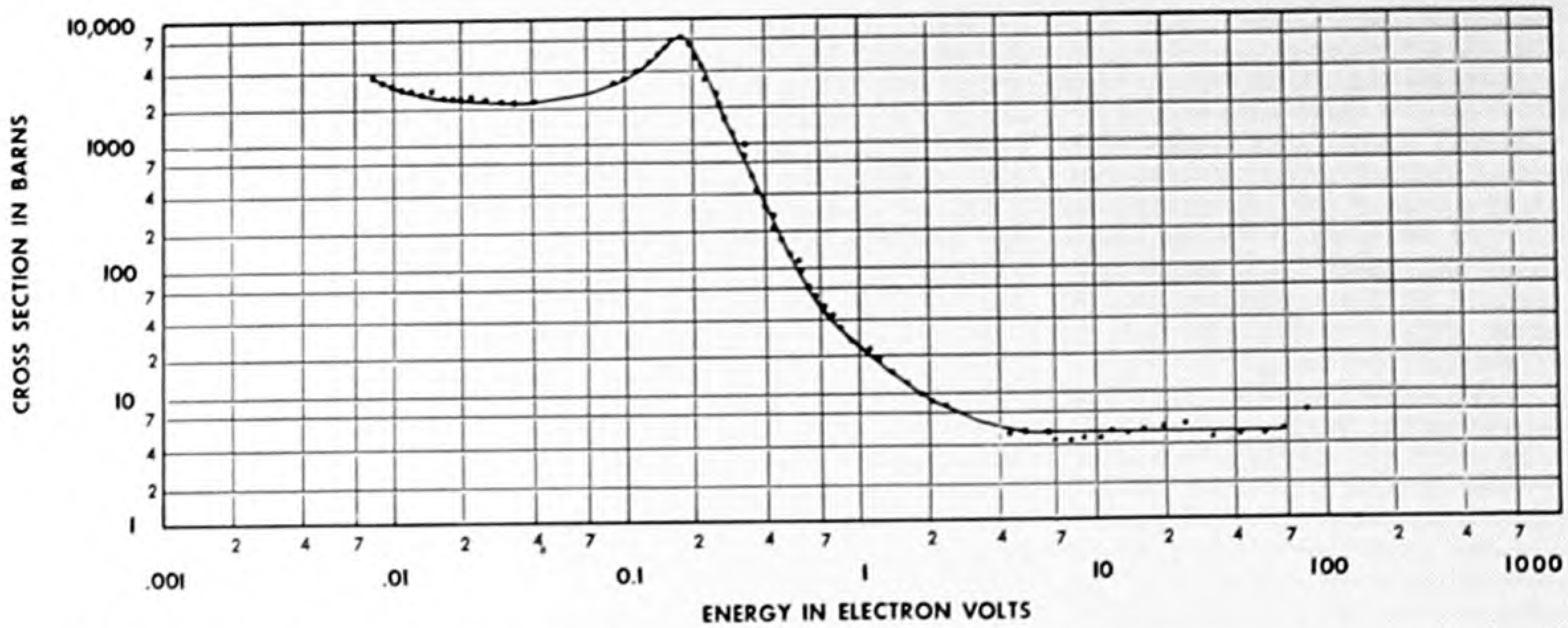
A slow neutron is the most efficient nuclear reagent known. A slow-moving proton or alpha particle is kept far away from a nucleus; the electrostatic repulsion a positively-charged nucleus exerts on a positively-charged proton or alpha particle keeps away all but those with the highest energy. Once a proton or alpha particle has slowed down, it wanders about with no prospect of getting near a nucleus, but instead picks up an electron or two and settles for the quieter life of a stable atom. On the other hand, the slower the neutron, the more likely it is to be captured. No neutron escapes eventual capture by some nucleus, except for the one out of every few million that spontaneously decays. For the physicist who wishes to control the fate of his neutrons the chief problem is one of economy—he must prevent the loss of neutrons, whether through random diffusion out of the apparatus in which he is working, or from capture in the extraneous material within that region. The latter part of the problem is solved by the use of carefully purified moderators so that the neutron will have little opportunity to combine with any unforeseen substance. The former part of the problem resolves into a question of size; to prevent neutrons from diffusing out of his working region the physicist makes the region ever larger. A laboratory where neutron work is done is characterized by large piles of graphite blocks or boxes of paraffin or even tanks of the costly heavy water. Such arrangements are the closest approximation to bottles for the most active of all reagents.

The Neutron as a Wave

Thus far we have spoken of the neutron as a particle, a hard round speck of stuff. But this is only one aspect of the neutron; indeed, it is only one aspect of all the atomic material of modern physics. Not only is the neutron a particle, it is also a wave. This subtle notion of the duality of matter is the fundamental

PARTICLE	CHARGE	MASS
ELECTRON	−	1
POSITRON	+	1
PROTON	+	1836
NEUTRON	0	1836
PI MESON	−	276
PI MESON	+	276
MU MESON	−	210
MU MESON	+	210
NEUTRAL MESON	0	271
V PARTICLE	−	?
V PARTICLE	+	?
V PARTICLE	0	2100 ?
NEUTRINO	0	0 ?

GLOSSARY of the fundamental particles of matter lists their electrical charge (+ or —) or lack of it (0). Also shown is the mass of the particles in terms of that of the electron.

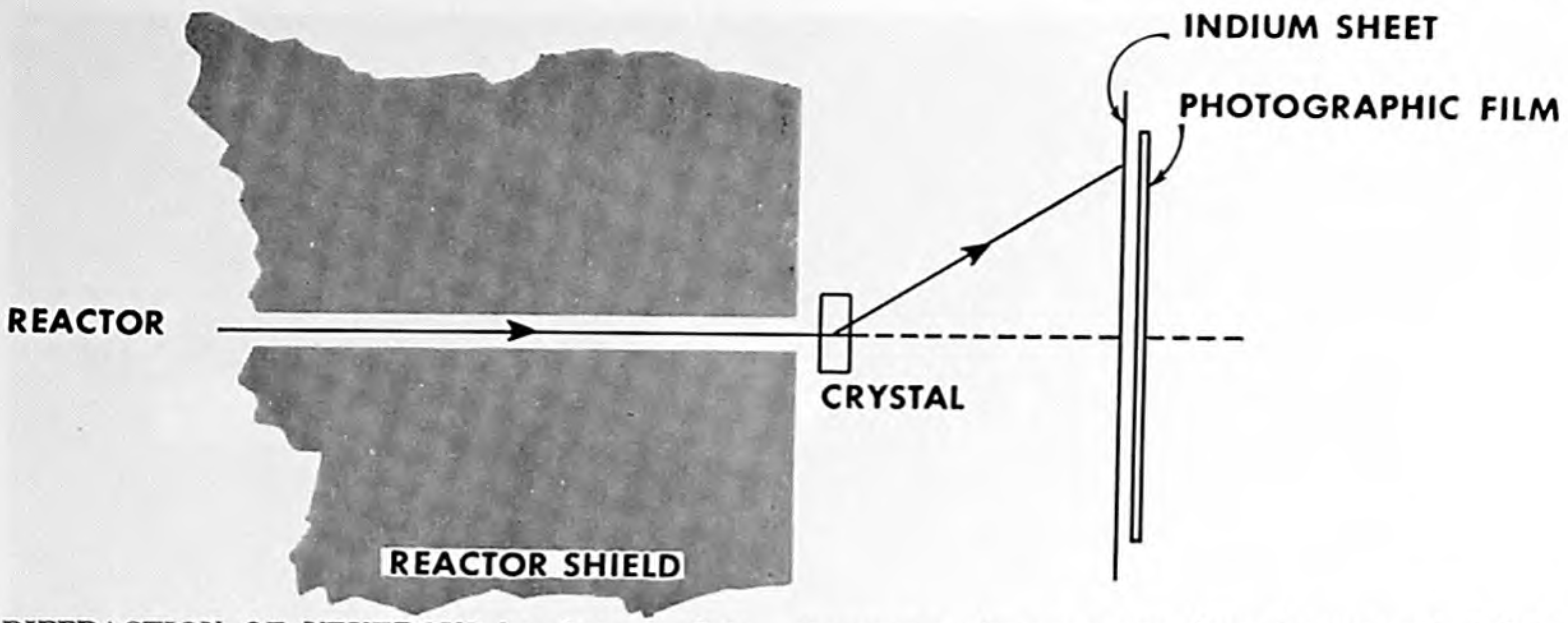


ABSORPTION OF NEUTRONS by the nuclei of the element cadmium is shown by this curve. Each point on the curve indicates the cross section, *i.e.*, the probability of capture, for a neutron of given energy. The "barn" is 10^{-24} square centimeters. The curve shows that cadmium is a very efficient absorber of slow neutrons.

principle of quantum theory, that indispensable language of the modern physicist. Perhaps the most spectacular aspect of the neutron as a wave is the phenomenon of nuclear resonance. Resonance is the familiar but always marvelous concomitant of any periodic disturbance. In its essentials nuclear resonance has much in common with a child on a swing. By gentle but carefully timed pushes, adjusted to the particular swing, the child can send the swing far above the bar from which it is suspended. What happens when a neutron wave encounters a nucleus? Of course the nucleus is a little sphere with three dimensions, and neutron waves encounter it from all sides. But for simplicity let us use a one-dimensional analogy; then the nucleus has simply a front and a back edge which lie in the path of the neutron wave. The analysis required to extend this picture more realistically to three dimensions we can leave to mathematics. Now when a wave strikes any obstacle, part of it goes on, but some of its energy is scattered back. In our one-dimensional case both the front and back edges of the nucleus send a small portion of the wave back in the direction whence it came. If it happens that the portion diffracted by the back edge returns to the front edge in such a way as to be precisely in step with the incoming wave, we have a condition of resonance. The incoming wave is reinforced by the diffracted wave. Stronger now, the incoming wave will again diffract at the back edge and contribute a portion of itself, still in step after a complete round trip across the nucleus, to the next incoming crest. If these waves are exactly in step they will reinforce indefinitely; the amplitude of the wave inside the nucleus will, after many internal traversals, become very large. If we use the quantum language to translate wave into particles we find we have been saying that captured neutrons are very likely to be found within the nucleus. The reinforcement requires that the time between successive incoming wave crests precisely match the time for the interior wave to complete the round trip across the nuclear obstacle. The incoming wavelength, which is fixed by the energy of the neutrons, will determine the entire course of events. For exactly the right wavelength, or neutron energy, the match will be perfect: the nucleus will voraciously absorb neutrons; for an energy just a trifle different, the nucleus will be almost transparent to neutrons. The resonance can be broad or narrow, just as a tuned radio circuit can have a sharp or broad response. The study of resonant energies and the breadth of their peaks comprises the study of all internal motions of the nucleus. The study is vigorously pursued by physicists over the whole range of neutron energies and over all available target nuclei. What is especially interesting for physicists who study nuclei with neutrons is the great effectiveness of slow neutrons. The fact that neutrons have no electric charge and thus experience no electrostatic repulsion means that they can interact strongly with nuclei even when very slow. A beam of thermal neutrons moving at about the speed of sound, which corresponds to a kinetic energy of only a fortieth of an electron volt, produces nuclear reaction in many materials much more easily than a beam of protons of millions of volts energy, traveling thousands of times faster. A doubling of the neutron energy represents a large and thus easily measurable percentage change, but corresponds to a small fraction of a volt energy increase, a variation which would be undetectably small against the million-volt energy of the charged-particle beam. An energy change of a few thousandths of a volt is enough to carry the neutrons across an entire nuclear resonance peak. This enormous and unexpected magnification of the first volt or so on the energy scale —the energy range occupied by slow neutrons—is perhaps the single greatest aid to the unraveling, still very much unfinished, of the structure of all but the lightest nuclei. It is a little out of our scope, but worth the digression, to point out that on the energy scale the position of just one of the millions of resonance levels which could be observed for a uranium nucleus is decisive for the success or failure of the slow-neutron chain reaction. Special devices that produce beams of neutrons of a single and carefully controlled energy are used for the study of nuclear resonance. These devices, which have been made in wide variety, depend on crystal effects or on careful mechanical or electrical timing of neutron flight to produce their single-energy beams. At somewhat higher energies the experimenter manages to use nuclear reactions for the generation of neutrons which are of such a kind that the neutron energy varies as he varies the energy of the charged-particle beam that produces the reaction. The whole neutron energy range from zero up to a few million volts has been studied with some care. The first important volt is now expanded into a thousand analyzed stretches; clearly for sheer bulk of information such a detailed study cannot be spread over a million volts.

Neutron Diffraction

The wavelength of thermal neutrons is several angstrom units, comparable to the distance between atomic nuclei in a crystal (an angstrom is a hundred millionth of a centimeter). The neutron wave can thus be diffracted by crystals.



DIFFRACTION OF NEUTRONS has been used to study crystals by C. G. Shull and E. O. Wollan of Oak Ridge National Laboratory. In their experiments neutrons escape from a reactor through a quarter-inch hole in its shielding. When the neutrons are diffracted by a crystal, they make spots on a photographic plate.

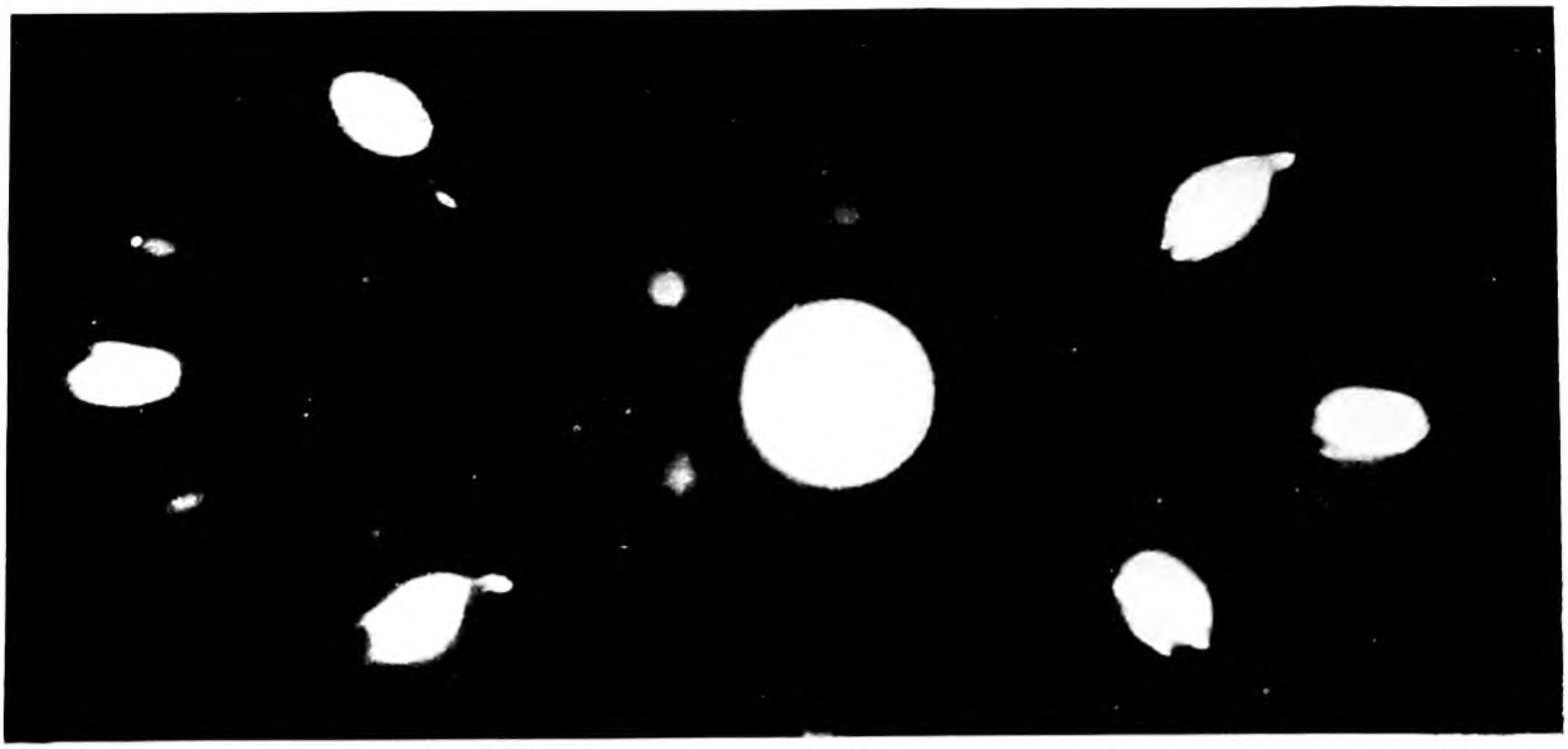
To clarify the notion of diffraction, hold under a lamp a microgroove phonograph record slanting away from your eye, and you will see a series of rainbow colors. This is the diffraction pattern made by light waves striking the grooves. The space between grooves is 50 times the wavelength of visible light, but this is fine enough to show visible, though rather crowded, patterns. For precision optical work, where clearly defined and widely separated diffraction patterns are essential, the physicist uses glass carefully ruled with some 10,000 parallel grooves to the inch. Each groove scatters the light that hits it, and the regularly spaced array of grooves works to reinforce the light scattered in a few special directions to form regular diffraction patterns. The sharp definition of these patterns depends on the precise spacing of the grooves. For the diffraction of X-rays, whose wavelength is thousands of times shorter than that of visible light, no man-made rulings can be fine enough. The regularly spaced atoms of crystals, however, make an excellent submicroscopic diffraction grating. The physicist has used X-ray diffraction techniques for decades to study crystal structure. The geometric arrangement of atoms within a crystal is made vivid by the famous Laue spots, the dots on a photographic plate made by X-rays diffracting from rows of electrons within the crystal. Now just as X-rays are scattered by the electrons of atoms in a crystal, neutrons are scattered by the nuclei of such atoms. The physicist today has added a neutron source and detector to the standard X-ray diffraction apparatus so that diffraction techniques may now be applied to the study of nuclear properties. The diffraction of slow neutrons is also useful to those who study crystal structure by diffraction. In ordinary X-ray diffraction pictures the very light atoms like hydrogen can hardly be detected, especially when they are found combined with heavy atoms in some crystal; the few electrons of the light atoms are masked by the many electrons of the heavy atoms. In neutron diffraction electron charges have no effect, and only the nuclei show. Light nuclei may be seen as well as heavy ones, often even better because their diffracting properties depend not on weight but on specific nuclear differences. Among the most recent studies made possible with this new technique has been one of the structure of the crystal ice; this confirmed previous indirect assumptions about the location of the hydrogen nuclei in a network of frozen water molecules.

Even more familiar than diffraction are refraction and reflection. All three are properties which depend on the fact that waves can be scattered. In diffraction the responsible scattering centers are individual, regularly-spaced obstacles, but in reflection or refraction the scattering is the statistical summation of the effects of all the atoms of the material involved. Visible light can be reflected from a glass or metal mirror at almost any angle; X-rays and neutrons can be reflected appreciably only at low angles. In some recent investigations neutrons have been reflected from the mirror surface of a liquid mixture of hydrocarbons, *i.e.*, compounds of hydrogen and carbon. This experiment has made possible a precise and complete measurement of the various interactions of the neutron with carbon and with the proton of hydrogen. Because of the comparatively simple "optical" nature of the experiment, such interactions can be studied by measuring the angles at which complete neutron reflection takes place. Once such interactions could be observed only by making measurements, never very accurate, of the passage of neutron beams—regarded as a stream of particles—through waxy filters. The neutron-hydrocarbon reflection studies, though essentially the studies only of neutron wave effects, throw a direct light on the behavior of the neutron as a particle.

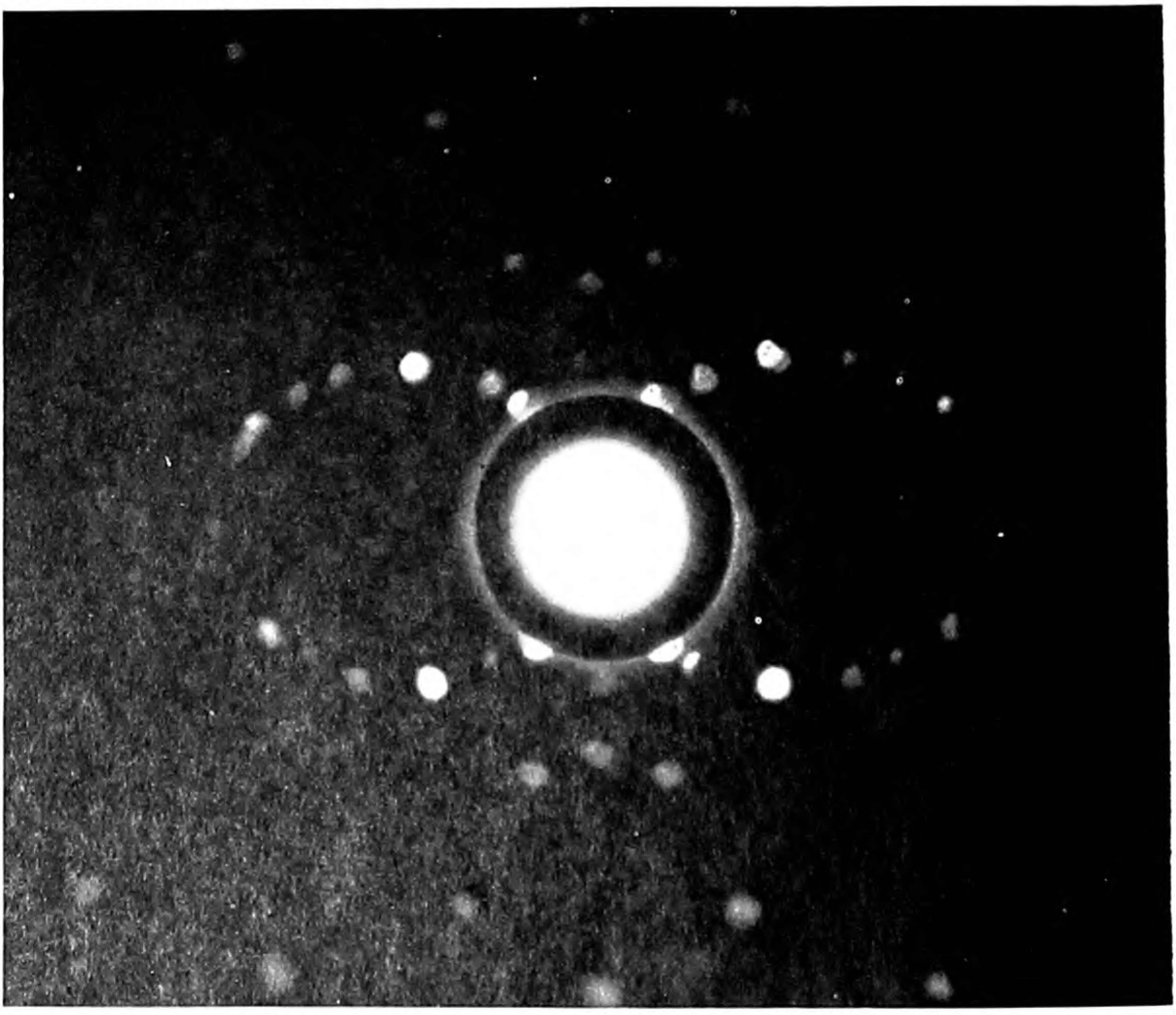
New Questions

Up to now we have discussed facts about the neutron, the hard core of facts which have been tested experimentally and are clearly understood. But in the 20 years since the neutron was discovered, a great many new questions have been raised. Some of them have been answered definitively, but many others remain unsolved. Let us consider these latter questions, along with most physicists. If we were to plot on a graph the growth of the physicist's understanding of the neutron, we would have a reasonably smooth ascending curve for the first 10 or 15 years, but during the past few years we would have a profusion of scattered points with no one clearly defined direction. Perhaps after another decade the dots will have formed a clear pattern, and the curve will continue to a peak of real understanding. At present, however, the status of the neutron as a fundamental particle is far from clear. What follows is a summary of current ideas: shaky speculation combined with firm experiment.

Is the neutron really neutral? Up to the time the neutron was discovered, all other particles known to have a mass also had an electric charge. The question of neutrality has been troubling physicists since the day neutrons were discovered. Charged particles are deflected by electric or magnetic fields; Chadwick himself tried in vain to deflect the neutron with an electric field. His successors have likewise failed to observe any de-



BERYLLIUM CRYSTAL is photographed by the technique depicted on the preceding page. The principal utility of neutron diffraction for crystal studies is to show the position of light atoms masked by heavier ones.



QUARTZ CRYSTAL makes a neutron diffraction pattern different from that of beryllium. Like X-ray diffraction patterns, neutron diffraction patterns can be mathematically read to locate the atoms in a crystal.

flection of even the slowest neutrons with the strongest electric fields. Charged particles knock electrons from atoms, leaving ions; no ionization has ever been noticed along the track of a neutron except when charged particles have recoiled as the result of a direct hit on a nucleus. The beta-decay experiments which showed that the neutron decomposes into a proton and electron—two particles of equal and opposite charge—confirmed all the previous assertions as to the over-all electrical neutrality of the neutron. It was found long ago, however, that the neutron has magnetic properties, that it lines up like a compass needle of a tiny but definite strength or "magnetic moment" in an experimental magnetic field. The deuteron, which, to repeat, consists of a neutron and a proton, was found to have a magnetic moment only about one-third that of the proton by itself. The proton was known to have almost three of the natural nuclear units of magnetic moment. The difference between the moment of the deuteron and that of the proton was accounted for by an assumed magnetic moment for the neutron of about minus two units. This assumption was confirmed by direct measurement in an experiment done just before the war. Just as iron filings can be guided by a magnet, so it was found that slow neutrons could be deflected from their course by passing them through magnetized iron. For the first time the effect of a magnetic field on a neutron was clearly observed. The minus sign for the neutron's magnetic moment indicates that the direction of the neutron's magnetic poles is mechanically opposite that of the proton's. One may visualize the neutron and the proton as tiny gyroscopes, each with a bar magnet as its axis. If a neutron and a proton are spinning in the same direction, their magnets will point in opposite directions.

It is very difficult to believe that the neutron can be magnetic and yet be associated with no electrical charge whatever. Recent experiments seem to indicate, in fact, that the neutron consists of a positively-charged core surrounded by a thin shell with an equal quantity of negative charge. Again to repeat, the neutron responds primarily to nuclear forces which reach only some 10^{-13} centimeters into space, while the distance from an atomic nucleus to one of its electrons may be 100,000 times as far. The neutron thus rarely hits an electron; at any reasonable distance the neutron appears entirely neutral, its interior positive charge canceling out its exterior negative one. If a neutron comes very close to an electron, so that the electron lies partly within the negatively-charged shell, the neutron should no longer appear entirely neutral. In the experiments done to test this hypothesis, slow neutrons were scattered from the atoms of xenon. Xenon is the heaviest and hence the most electron-rich of all the inert gases. These gases form no chemical compounds whatever. The theory of atomic structure can explain this property by the circumstance that the electrons in such gases are arranged so that their magnetic effects cancel completely. The magnetic moments of the electrons are paired off, each moment canceled by one aligned in the opposite direction, so that there is no possible net magnetic interaction between a xenon atom and any external magnet such as a neutron. Now the nuclei of atoms scatter very slow neutrons uniformly in all directions, essentially because nuclear interactions are localized at the very center of the atom. If the atoms of xenon scattered neutrons unevenly in various directions, it must be because there is some interaction between the electrons of xenon and the neutrons. But this could not be magnetic: the magnetic forces canceled. Yet each electron acted as a weak scattering center, its effect being about a thirty thousandth as strong as that of the nucleus. The net result was not entirely unambiguous, but together with related experiments it indicates that the neutron does indeed interact, though very slightly, with the electron's charge. There is no doubt that the neutron has a charge-bearing structure, though it is neutral overall.

Neutrons in Nature

In the earth and the air around us, indeed in our very bodies, there is always a perceptible background of neutrons. The radium-beryllium source used to produce neutrons in the laboratory has its counterpart in nature, both deep in the earth and high in the air. Rocks, though almost entirely composed of light elements, generally contain small quantities of heavy radioactive elements. Occasionally alpha particles released by these atoms of the radioactive uranium family strike against and react with the nuclei of the abundant silicon or magnesium atoms within the rock, just as they strike the nuclei of beryllium in a laboratory source. The result of these collisions within the rocks is a small, steady glow of neutrons, a few of which inevitably escape into the air. An even richer source of natural neutrons operates high in the atmosphere, mostly above six miles. There the harsh rain of primary cosmic rays beats down on the atoms of the atmospheric gases. When one of these high-energy cosmic particles encounters the nucleus of an atom, the nucleus explodes into fragments; the pronged image left by such an explosion in a cloud chamber or photographic plate is called a "star." Each of the many rays of the star corresponds to the track of a charged particle released in the explosion. For every particle leaving a visible track, however, usually one neutron goes off leaving no track at all; some neutrons travel through the air for miles before being slowed and captured. The production of neutrons in rocks and in the upper air is a slow process compared to the wholesale production of the laboratory. But as far as neutron energy is concerned, the neutron trickle from rocks and the neutron bursts of cosmic rays are roughly two extremes. Neutrons set free in the laboratory run the gamut of energies in between, sometimes indeed a bit lower than those made in rocks, but not yet and perhaps never as high as those made in some cosmic ray explosions.

Since the discovery of the neutron, and largely with its help, a tremendous lore about the transmutations of nuclei has developed. In place of the chemist's familiar periodic chart of the elements, the wall of a nuclear laboratory anywhere displays a nuclear isotope chart which plots all the known stable and unstable nuclei. This chart is a compact summary of facts the detailed description and analysis of which would fill a good many volumes. Each nucleus has its own square, whose position on the chart is determined by the number of protons and neutrons in the nucleus. Within each square is packed a vast amount of abbreviated information. From this chart one can tell which nuclei are stable and how abundantly these stable isotopes occur in nature; the half-lives of unstable elements and the particles they emit; and what nuclear reactions occur when the particles are captured. *The Physical Review*, like every other journal of physics, is constantly filled with papers reporting the discovery of new isotopes and new properties of known isotopes, so that a nuclear chart is barely printed before it needs revision. The crowded nuclear chart is, of course, merely a symbol for the whole study of nuclear reactions, vast enough to constitute an independent branch of physics. The experimenter has measured and described in detail as much as he can of the events which occur in a long series of possible nuclear reactions. If, for example, he bombards boron 10 with neutrons of a given energy to produce lithium 7 and alpha particles, he counts the number of alpha particles and lithium nuclei and then calculates the probability for the reaction to occur. The boron hit by a neutron generally produces a gamma ray in addition to lithium 7 and an alpha particle of less energy than before. He counts the number of gamma rays also and determines the degree of competition which this alternative to the first reaction represents. He uses his instruments to observe the boron target from many different angles and notes the rate of alpha particle production at each angle. He varies the energy of his incoming neutrons and studies the

effect of such variation on the yield of all the reaction products. The study of the neutron-boron 10 reaction outlined sketchily here has been repeated in greater or less detail for literally hundreds of different reactions. All the facts and figures assembled in these many experiments have tended to confirm the fundamental picture upon which the physicist builds: the nucleus is a tightly-bound collection of protons and neutrons whose structure he intends to unravel.

Cosmic Rays

The climax of this intensive study of nuclear reactions is the study of the really high-energy events such as the biggest cosmic ray explosions and the reactions recently produced by the high-energy particle accelerators of the post-war years. By exposing photographic plates to cosmic rays the physicist can display in minute detail the complicated pattern of an event which occurs to a single atom, but with an enormously concentrated energy in a thousand millionth of a second. He is able to do this because each charged particle released in the explosion radically disturbs each tiny crystal of the photographic emulsion through which it passes, in just the same way that light disturbs such crystals in ordinary photography. Because the sensitive emulsion of the photographic plate is a layer no thicker than a sheet of heavy paper, the path of a particle through it is very short, unless by happy accident it is closely parallel to the plane of the emulsion. Any detailed examination of the individual crystal grains in the emulsion requires a microscope. Seen through a microscope the track of a fast proton, say, appears to be an irregular row of tiny dark dots, each dot representing what was once a translucent grain of silver bromide, so affected by the rude passage of a charged particle that it turned to a black metallic speck of silver when the developer solution reached it. Because they do not leave the tracks of charged particles, neutrons cannot be seen directly, but their presence in cosmic rays is made abundantly clear by their other effects. After a painstaking examination of such a cosmic ray photograph, the physicist is able to describe an explosion in detail. He knows which particle set off the explosion, since he can learn its mass, charge and energy by measuring such quantities as the grain spacing and the deviations of its tortuous path through the emulsion. He measures in the same way the tracks of all the various particles produced by the explosion. From these studies of nuclear explosions have emerged a whole series of facts which play an important part in our knowledge of nuclear reactions.

The most exciting find resulting from the observation of cosmic ray explosions in photographic emulsions was the "pi meson," a new particle which interacts with neutrons and protons. Pi mesons have a mass about one-sixth that of a proton. Three kinds occur: one with a positive charge, one with a negative charge and one with no charge at all. Such mesons are transient; when free they last only a hundred millionth of a second and usually travel only a few feet in air. Although they are intrinsically so different, one may draw an analogy between the role of mesons and that of light "particles," or quanta. When electrons are ripped from an atom, the energy and momentum of the electromagnetic field of force which holds the charged particles of the atom together are partly released in the form of light quanta which travel out into space indefinitely, far from the disturbed atoms. We say that the light is radiated. Now if we consider the nucleus to be a collection of neutrons and protons held together by nuclear forces, we may by analogy expect that when a nucleus is disturbed and ripped apart, the energy released appears partly in the form of mesons, which play the same role for nuclear forces as light quanta do for the electromagnetic binding forces of the atom. In each case the very action of rearranging an existing structure, either atom or nucleus, creates something new which radiates afar. In one case we have light quanta, which have no charge and no mass, and in the other case free mesons, which have both charge and mass. This notion of a complete analogy between light quanta and mesons is an extension of the idea of treating the nucleus as a kind of condensed atom. If we look at the nucleus this way, the forces that hold it together involve a kind of incessant exchange of mesons between the nuclear particles. The idea was first clearly expressed by the Japanese physicist Hideki Yukawa about 15 years ago, and it has remained a major guide for such investigations ever since. For some time the study of mesons was confined to their analysis in cosmic rays, but new particle accelerators have been able for three years to make mesons artificially. Meson research is now going forward vigorously in many laboratories. Several different types of mesons have been discovered in a very few years, but of these the pi meson just now appears to be most closely related to nuclear forces. It is the origin of these forces that remains unclear.

The Crisis of Physics

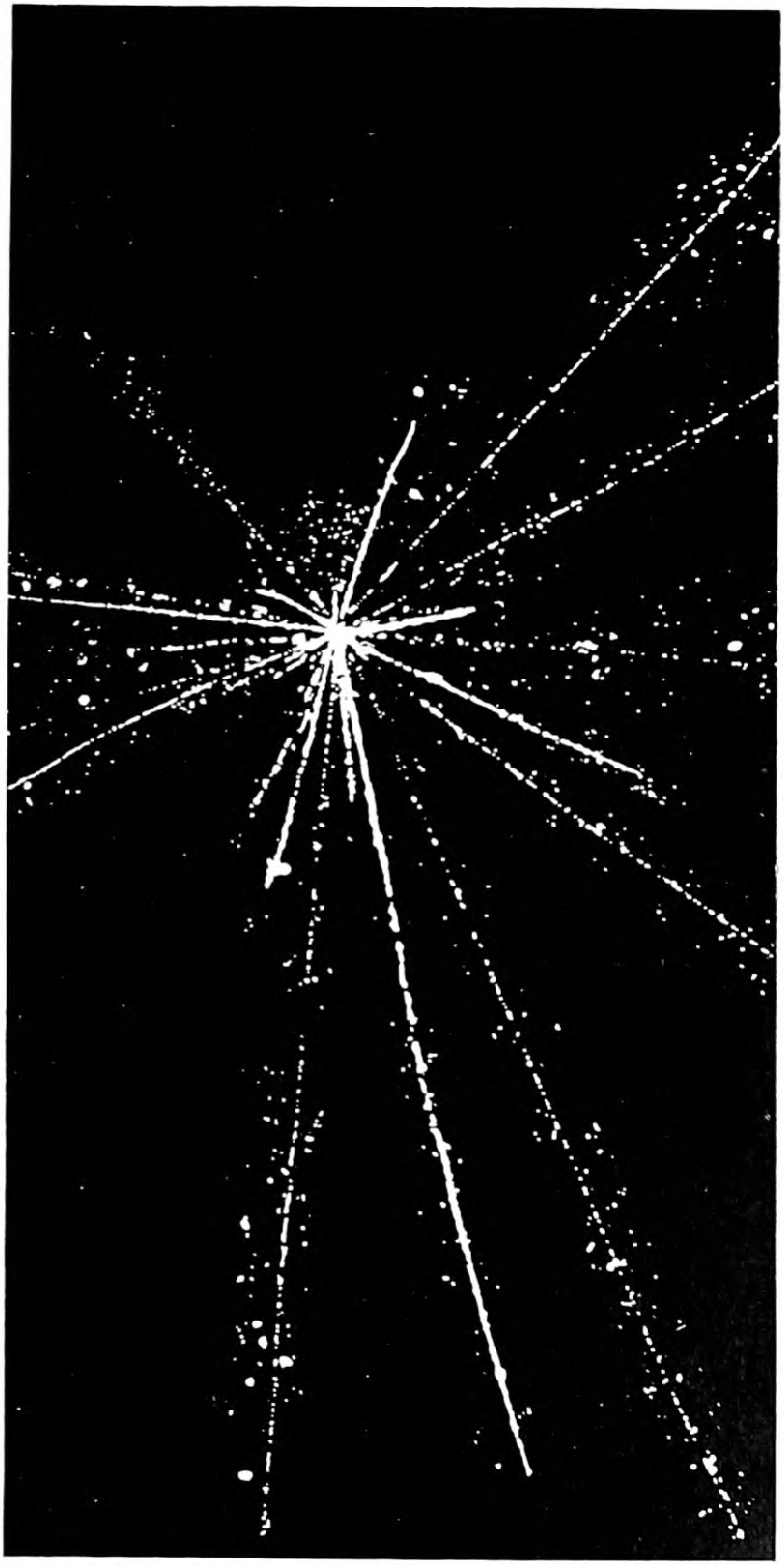
The crisis in physics centers about difficulties in the concept of fundamental particles. Physicists believe that protons and neutrons are fairly permanent constituents of nuclear matter, and in this sense they are fundamental. In no nuclear reaction at present-day energies does the total number of neutrons and protons change. The fundamental problem is still *why* these fundamental particles act as they do. Physicists have a reasonably clear and detailed picture, as we have seen, of *how* the neutron and proton act, and the experimenter can to a large extent control their interactions. So far so good. But the theory lags; real understanding is scant. Consider the electron. While the physicist cannot predict the charge or the mass of the electron, he can, given these numbers, fully account to six decimals for the magnetic moment of the electron. He can calculate in detail at any energy the interaction of electrons with each other and with light. This gives some confidence that the electron, whose properties are described so completely and so economically by the equations of the English physicist P. A. M. Dirac and the recent brilliant extensions of electromagnetic theory, is in a practical sense a simple and fundamental particle. Given a very few of its properties, the rest follow from the theories in rich detail. But as we have seen, this is not so for the nuclear particles, the proton and the neutron. The essential feature which seems to distinguish the cases is the strong interaction between the proton and neutron on the one hand and the meson on the other. We cannot even approximately regard the neutron or the proton as a well-defined and stable fundamental particle. On the contrary, these particles must be described as regions of interaction. A neutron must be pictured as a "bare" nuclear particle around which mesons steadily form and disappear. These are genuinely transient mesons; it is impossible to observe them as real particles, for the very process of observation disturbs the whole system enough to release them as free mesons, no longer bound. A neutron that became briefly a proton and a negative pi meson, or several mesons with net electrical charge of zero, and then recombined, would account graphically for some of the puzzling properties of this "fundamental" particle. The tentative picture is not unattractive, but it is quantitatively incorrect and generally intractable.

This rough working hypothesis is reinforced, however, by the observation that a neutron and proton may change places during a collision. A fast neutron goes by a proton and a fast proton comes out traveling in almost the same direction, while a neutron recoils rather slowly. This particular phenomenon occurs too often to be accounted for by a direct hit, which is after all a relatively rare occurrence; it must be happening even when the neutron merely passes fairly near the proton. Such an event could be accounted for by a transfer of a charged meson between the neutron and the proton. There may be many such exchanges. There is also some evidence, convincing but indirect, that electrical currents other than those accounted for by the motion of the proton flow in the nucleus.

The existence of nuclear forces of this type implies the easy transfer of charge between the proton and the neutron. The interaction of the many protons and neutrons in the nucleus under conditions which allow the flow of mesonic currents means that the identity of a nuclear particle is not fixed. The total charge remains immutable, but just which nuclear particle is a proton and which a neutron is a question that must be answered differently some 10^{20} times per second. The nucleus is no static realm of wandering planets, however pretty a picture this makes for the textbooks.

The Limits of Understanding

Our ignorance of the neutron curiously emphasizes the great strides that physicists have made in their understanding of the nucleus as a whole. They have classified a great number of nuclear reactions. They understand fairly well those reactions which occur in the upper atmosphere through the agency of cosmic rays and those induced by the beams of accelerating machines. They have learned how to manage and predict nuclear reactions, from those of the simple radium-beryllium sources to those of the fission processes of the uranium and plutonium bombs. They now have a considerable insight into the complex structure of nuclei. But physicists have only a confused and fleeting knowledge of the real nature of the parts which comprise that structure. The neutron and the proton can no longer be considered the fundamental and immutable building blocks of matter. The current list of "fundamental" particles now includes, in addition to the proton and neutron, the electron, the positron, the mu meson, the pi meson, the recently discovered V particles, the neutrino and the photon, or light quantum. Not all of these can be fundamental in any real sense of the word. The whole concept of matter must be extended somehow to include the different aspects of these particles in some unifying theory. Probably the real fault lies with the idea that one can separate out a fundamental particle. Some view of the joint and mutual interaction of the many "fields" of the particles whose tracks are so patiently and beautifully exhibited is perhaps the direction of the answer. But as in so much of physics the power of understanding is great but sharply and paradoxically circumscribed. The physicist controls the chain reaction, he catalogues the nuclear levels by thousands, he lists the artificially produced isotopes and their properties in grand array. But the inwardness of the neutron lies for a while beyond his understanding.



COSMIC RAY "STAR" was recorded in photographic emulsion by physicists of the National Research Council of Canada. The star represents the explosion of a silver or bromine nucleus struck by a high-energy cosmic ray.

The Authors

PHILIP and EMILY MORRISON are a husband-and-wife team who have collaborated on several previous articles for SCIENTIFIC AMERICAN. Morrison is associate professor of physics at Cornell University, where he has done pioneer work applying physical theory in a number of fields, including microbiology. He graduated from the Carnegie Institute of Technology in 1936, then studied theoretical physics under J. Robert Oppenheimer at the University of California, where he received his doctorate in 1940. When World War II broke out, Morrison left a lectureship at the University of Illinois to join the Metallurgical Laboratory of the University of Chicago, and later became a group leader at the Los Alamos Laboratory of the Manhattan District. Morrison joined the faculty at Cornell in 1946.

Bibliography

WHY SMASH ATOMS? Arthur K. Solomon. Harvard University Press, 1946.

THE "PARTICLES" OF MODERN PHYSICS. James Docking Stranathan. The Blakiston Company, 1945.

SCIENTIFIC AMERICAN August 1958, Vol. 199, No. 2, pp. 58-66 OFFPRINT 233

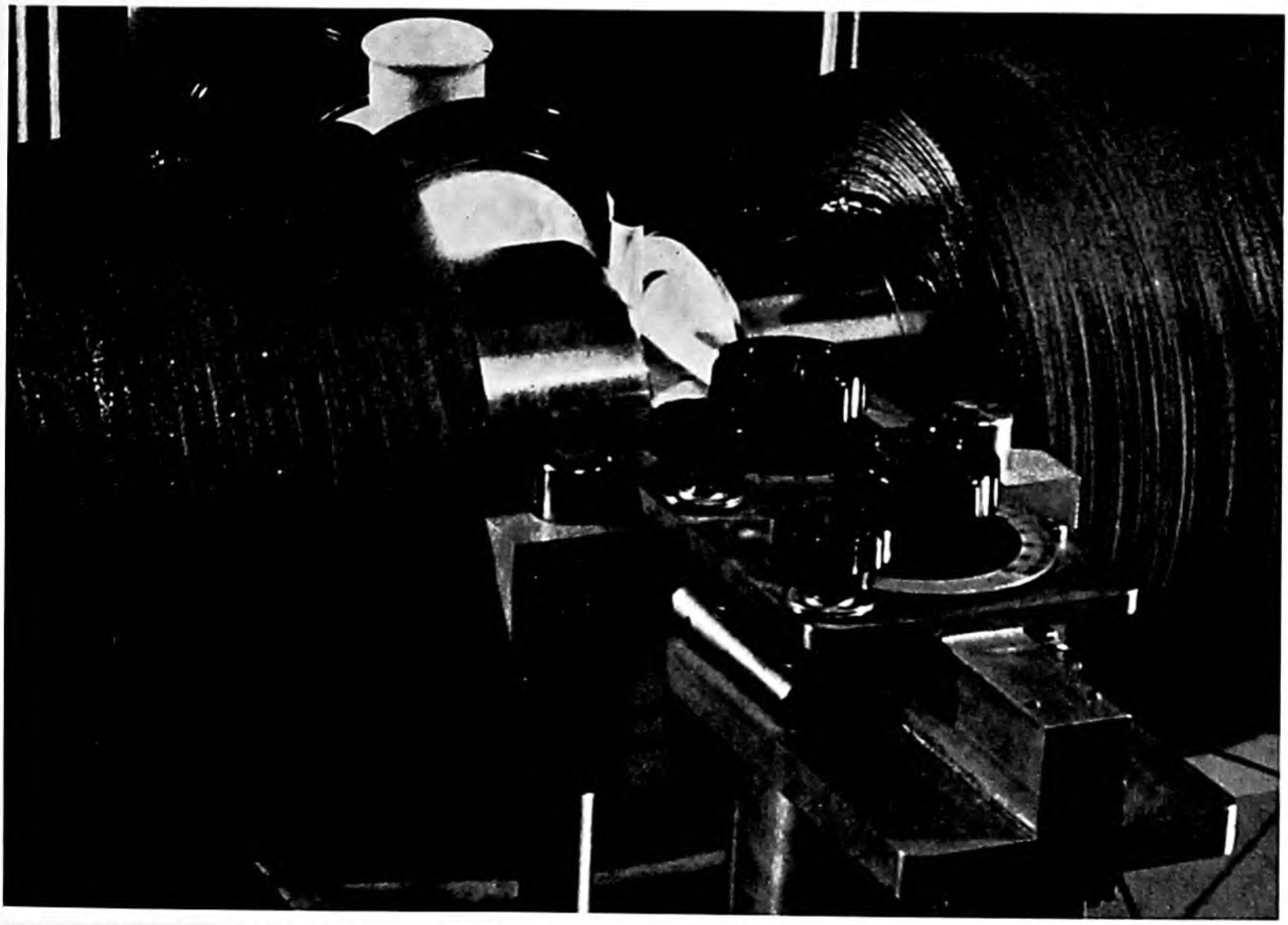
MAGNETIC RESONANCE

by George E. Pake

Atomic nuclei and electrons, which spin on their axes like tops, can be tipped over by magnetic fields. The technique furnishes much information about complex molecules and chemical reactions.

Early in 1946 Edward Purcell at Harvard University and Felix Bloch at Stanford University announced, almost simultaneously, an interesting discovery in physics. They had found a way to tune in on the magnetic fields of the spinning nuclei of atoms. The work was important enough to win the 1952 Nobel prize in physics for Purcell and Bloch.

It is doubtful that many people outside the field of nuclear physics were much excited by or even took notice of these experiments at the time they were announced. But by now the phenomenon in question, called magnetic resonance, has become a matter of very wide interest indeed. Scientists in various distantly separated lines of work—geologists, chemists, biologists—are, if anything, even more excited about the discovery



MAGNETIC-RESONANCE EXPERIMENT on electrons in chlorophyll demonstrates that photosynthesis involves free radicals. The sample of chlorophyll is in the upright tube in the center, between the poles of the magnet. The light source which illuminates sample is behind magnet. This photograph was made in the laboratory of Barry Commoner at Washington University in Saint Louis.

than nuclear physicists. For it has led unexpectedly to the development of a sensitive tool useful for a multitude of purposes, from prospecting for minerals in the earth to analyzing the chemistry of living organisms.

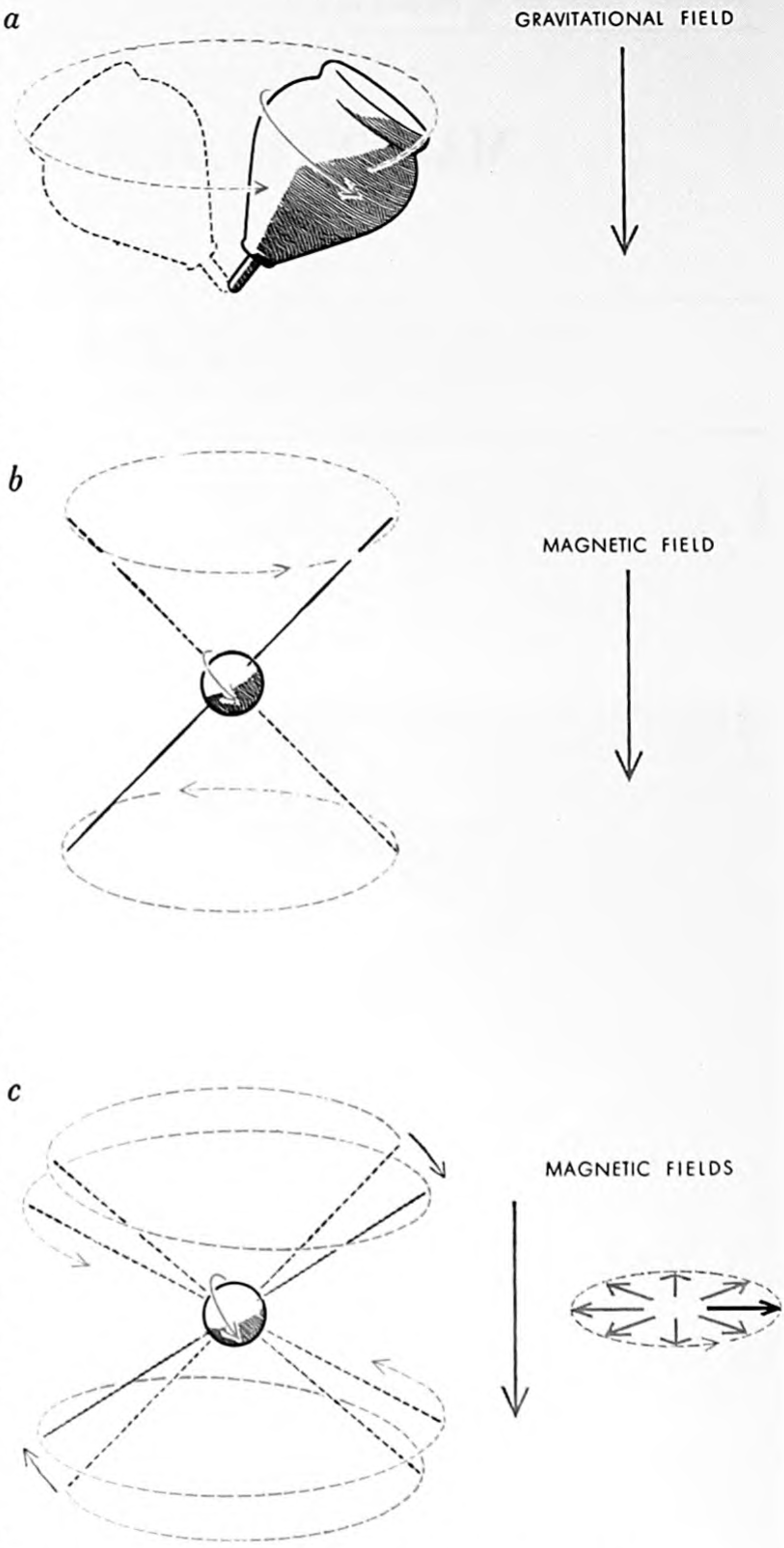
We must start this story by giving thought, as Purcell and Bloch did, to the spinning behavior of the proton. Like the rotating earth, a proton constantly spins around its axis (as do all the other elementary particles of the atom). It spins eternally with a certain momentum, reckoned in the proton's case as one-half unit. Since the proton carries an electric charge, its spin generates a magnetic field; in other words, the proton is a tiny magnet. Now we can ask ourselves the intriguing question: What will happen if we apply an outside magnetic field to this little magnet? Purcell and Bloch fell to wondering (independently) whether it would be possible to use a magnetic field to manipulate the spinning protons in a piece of matter: say, flip them over so that their north and south poles were reversed.

Here it is useful to think of an ordinary spinning top or a gyroscope. As everyone knows, a gyroscope tipped from the vertical does not fall down; instead its upper end circles slowly (precesses) around the vertical. That is to say, the downward gravitational pull of the earth acts to swing the axis of the spinning gyroscope around rather than to tip it further. Similarly, if we apply magnetic force to a spinning proton it will make the proton precess, not tip over. Brute force, in the form of stepping up the strength of the magnetic field, will avail us nothing: it will merely cause the proton to precess faster.

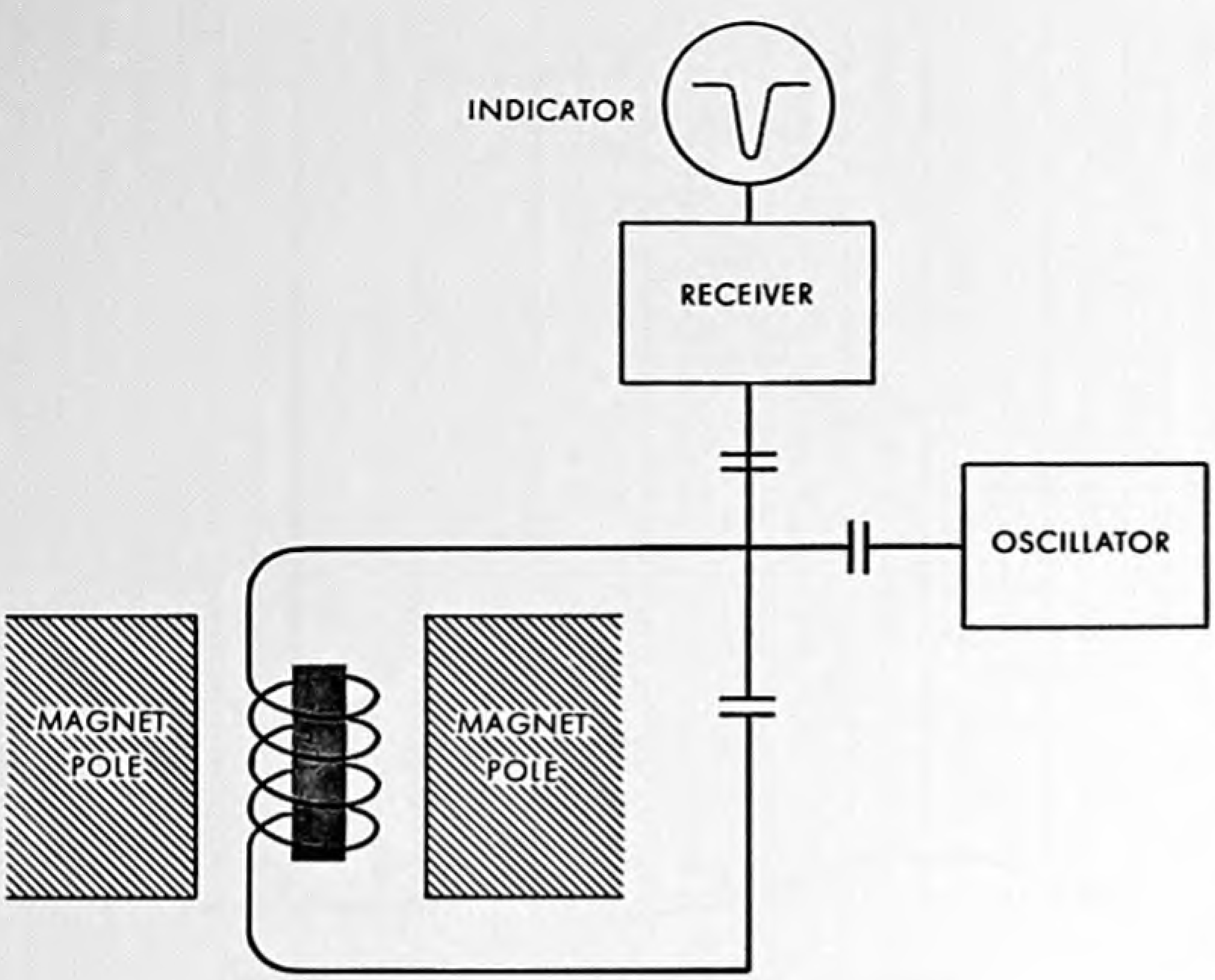
But there is a way to outwit the spinning particles. Suppose we apply a second magnetic field at right angles to the main field. Theory says that if we make the second field rotate around the first (by means of an alternating electric current in a coil), and if we time the rotation so that it coincides exactly with the rate of the proton's precession, we should be able to tip the proton over. In short, with proper tuning (at radio frequencies) the feat may be achieved by a magnetic-resonance effect.

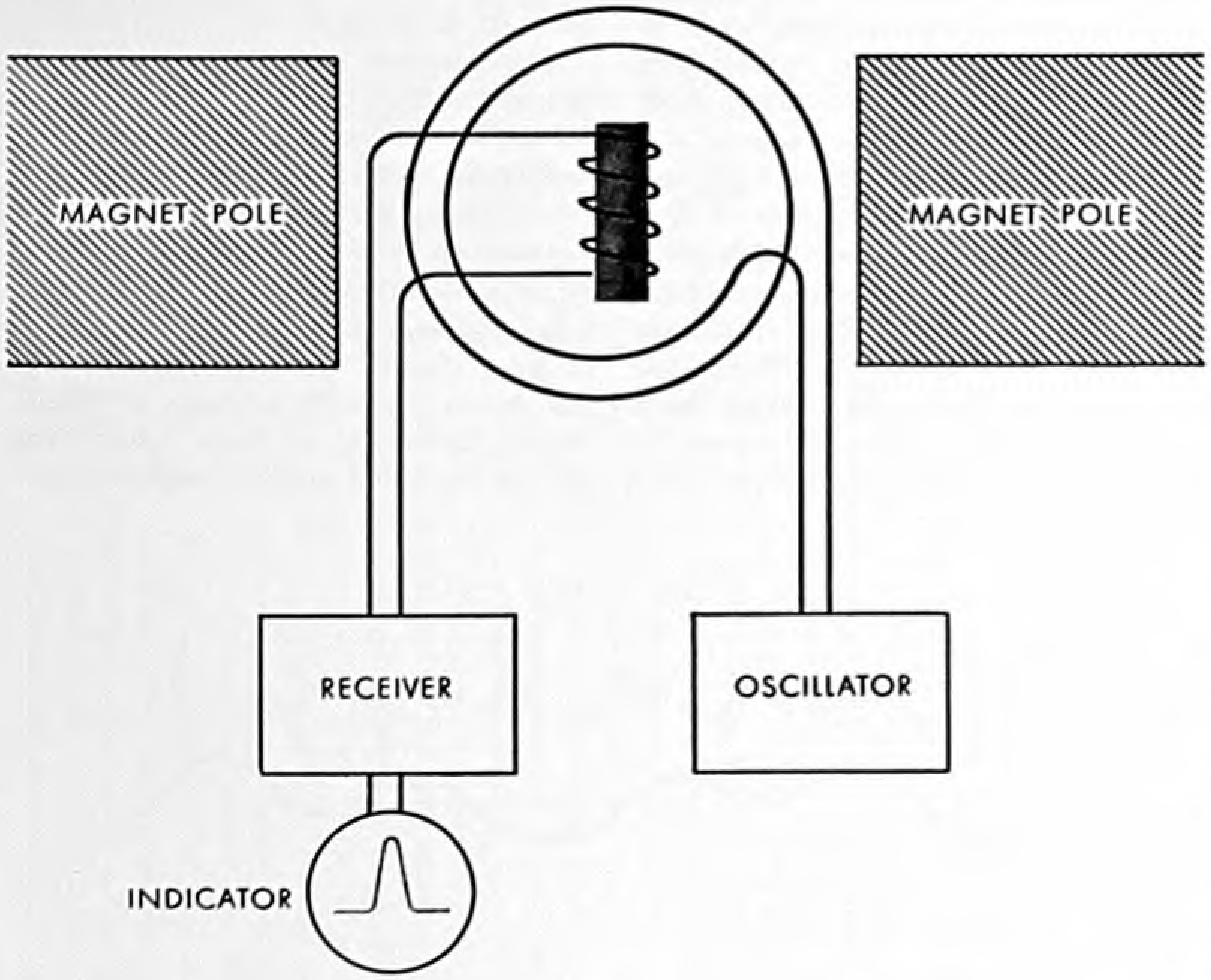
A New Spectroscopy

The problem, then, was to find the resonance frequency and to detect the effect on the protons. Purcell and Bloch employed different methods to look for the effect. In Purcell's apparatus the sample of matter was placed between



GYROSCOPIC ACTION of a spinning top or particle causes it to revolve or precess around a vertical field tending to tip it over. Particle can be tipped, as shown in the diagram at the bottom, by applying in addition a rotating horizontal magnetic field (*right*), which revolves at exactly the same rate as the axis of the precessing particle in the vertical magnetic field.





NUCLEAR-RESONANCE DEVICES designed by Edward Purcell and Felix Bloch are diagrammed schematically. In Purcell's arrangement *(top)* a single coil passes radio waves from oscillator into sample *(gray bar)* and feeds into receiver. Resonance is indicated by a dip in received energy. In Bloch's apparatus *(bottom)* a separate coil picks up energy from tipping particles. Resonance shows up as an increase in energy reaching the receiver.

the two poles of a magnet and surrounded with a coil which produced the second, rotating field [*see upper diagram at left*]. When the frequency was just right, energy passing along the coil was absorbed by protons in the sample of matter as they flipped over; this absorption of energy was recorded by a sudden dip in the strength of the signal reaching a radio receiver. Bloch's group, on the other hand, devised an instrument which recorded the event by induction of a voltage. As the protons flipped over, the motion of their magnetic fields induced a voltage in a second coil, and this signal was registered on an oscilloscope [*see lower diagram at left*].

The electron, like the proton, is a charged particle; it, too, spins and has a magnetic field—far stronger than the proton's, because it spins much faster. The electron also exhibits magnetic resonance. Since it is a stronger magnet and much lighter than the proton, it precesses much more rapidly in a given magnetic field. Whereas the proton is probed with radio waves in the range of a few megacycles per second (near the frequencies of ordinary home radio), for electrons the frequencies employed are in the microwave range, around 10,000 megacycles per second.

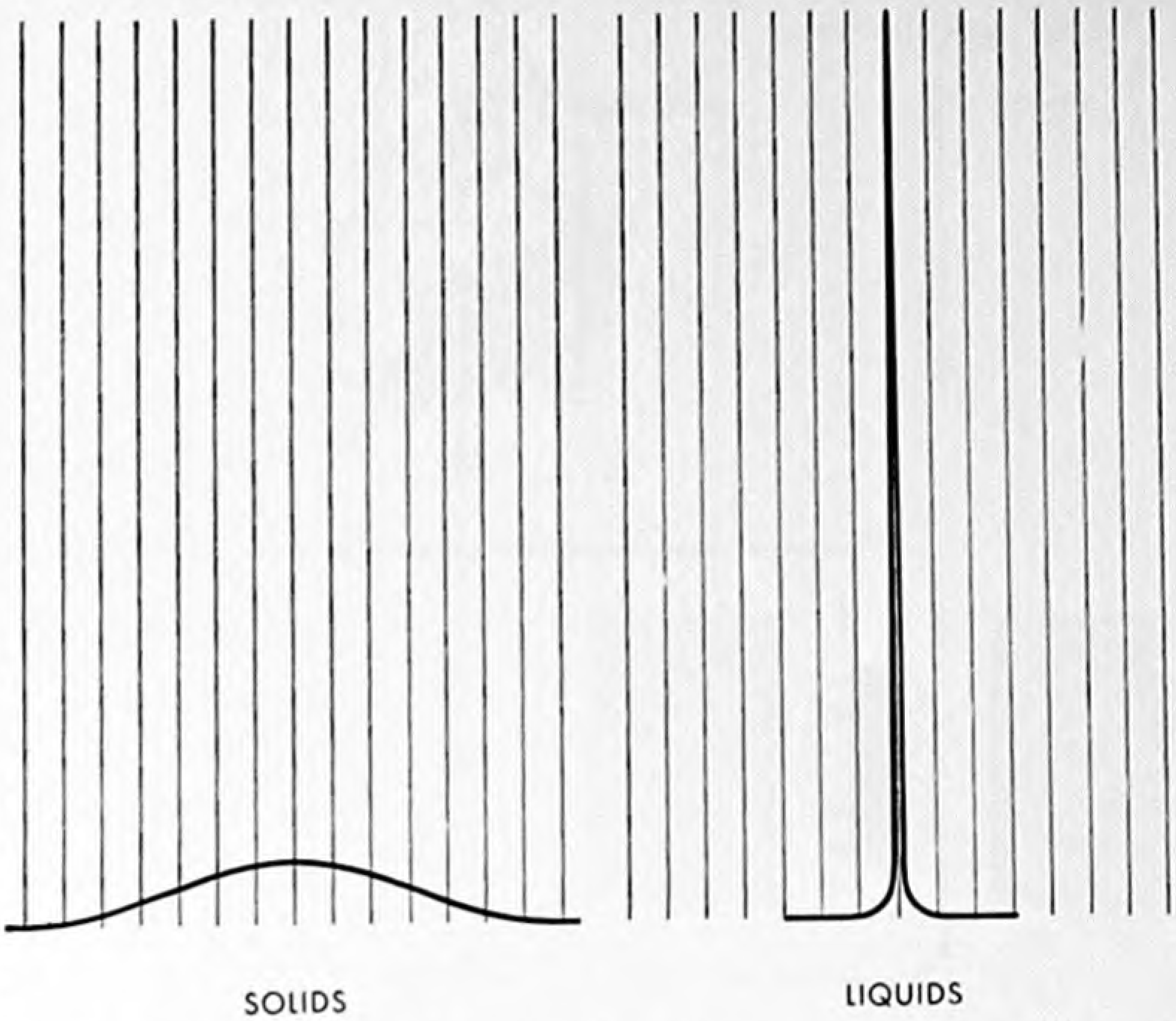
The magnetic resonance of protons and electrons makes it possible to learn many things about atomic nuclei, atoms and molecules. The magnetic probe amounts, in effect, to a new kind of spectroscopy. This brings us to the varied uses of the discovery in chemistry and biology.

The Structure of Molecules

Let us look, for example, at an organic compound such as cyclohexane (C_6H_{12}). We shall examine its resonance spectrum by means of nuclear magnetic resonance (NMR). We measure the substance's resonance in terms of the precession speed at which the nuclei flip over. That is, instead of tuning in to the resonance by varying the frequency of the rotating magnetic field, we use a fixed frequency and vary the strength of the main magnetic field, which controls the precession speed; when the nuclei flip, the resonance reading is taken as the strength of the magnetic field at that point (measured in gauss or oersteds).

When we subject a sample of cyclohexane at room temperature to the proper magnetic field, we get a sharp resonance reading. The compound's hydrogen nuclei flip simultaneously at a certain field strength (the carbon atoms are not

affected). Now the sharpness of this response has to be explained. In any collection of atoms we must reckon not only with the applied magnetic field but also with the little nuclear magnets of the atoms themselves. Each nucleus is subject to the magnets of its neighbors as well as to the applied field. The total magnetic field acting on a nucleus must vary from place to place in the material, because of the varying orientation of the nuclear magnets that happen to surround it. As a result of these variations, the responses to the external field should vary: it should take a slightly stronger applied field to flip some nuclei than others. This means that the resonance range for the whole group of nuclei should be a broad band rather than a sharp line. However, in a liquid the local variations are so short-lived, because of the rapid random motions and mixing of the molecules, that in effect all the nuclei are subject to about the same average field. Thus cyclohexane, a liquid at room temperature, gives a sharp magnetic-resonance line.

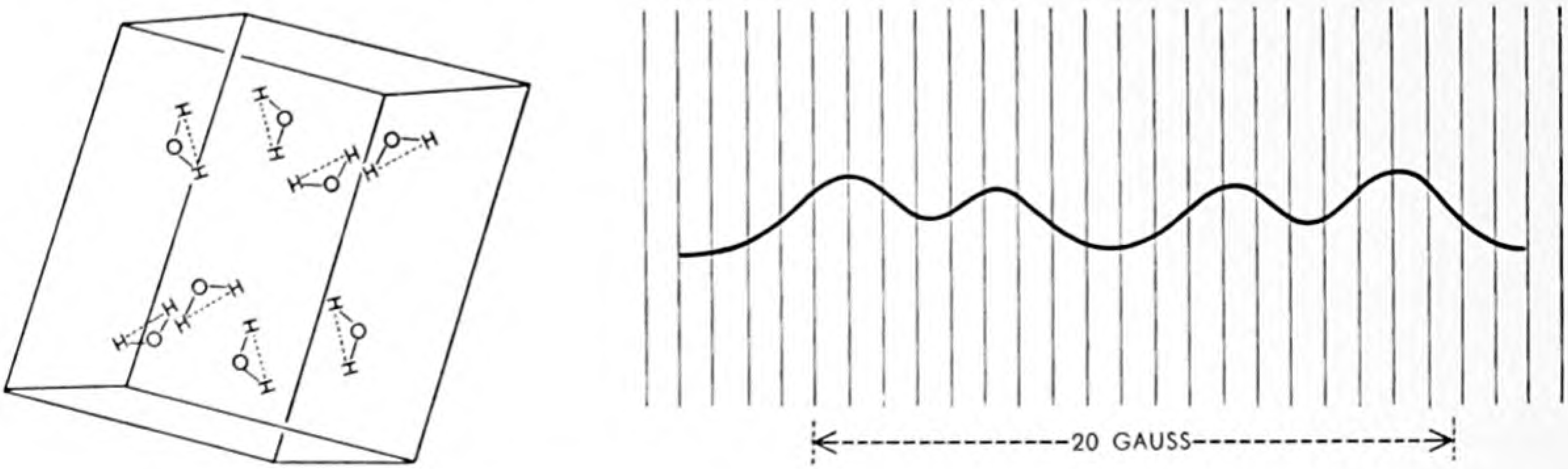


NUCLEAR-RESONANCE CURVES are broad for solids, narrow for liquids. The difference is actually greater than indicated here: the ratio of widths may be as great as 100,000 to 1.

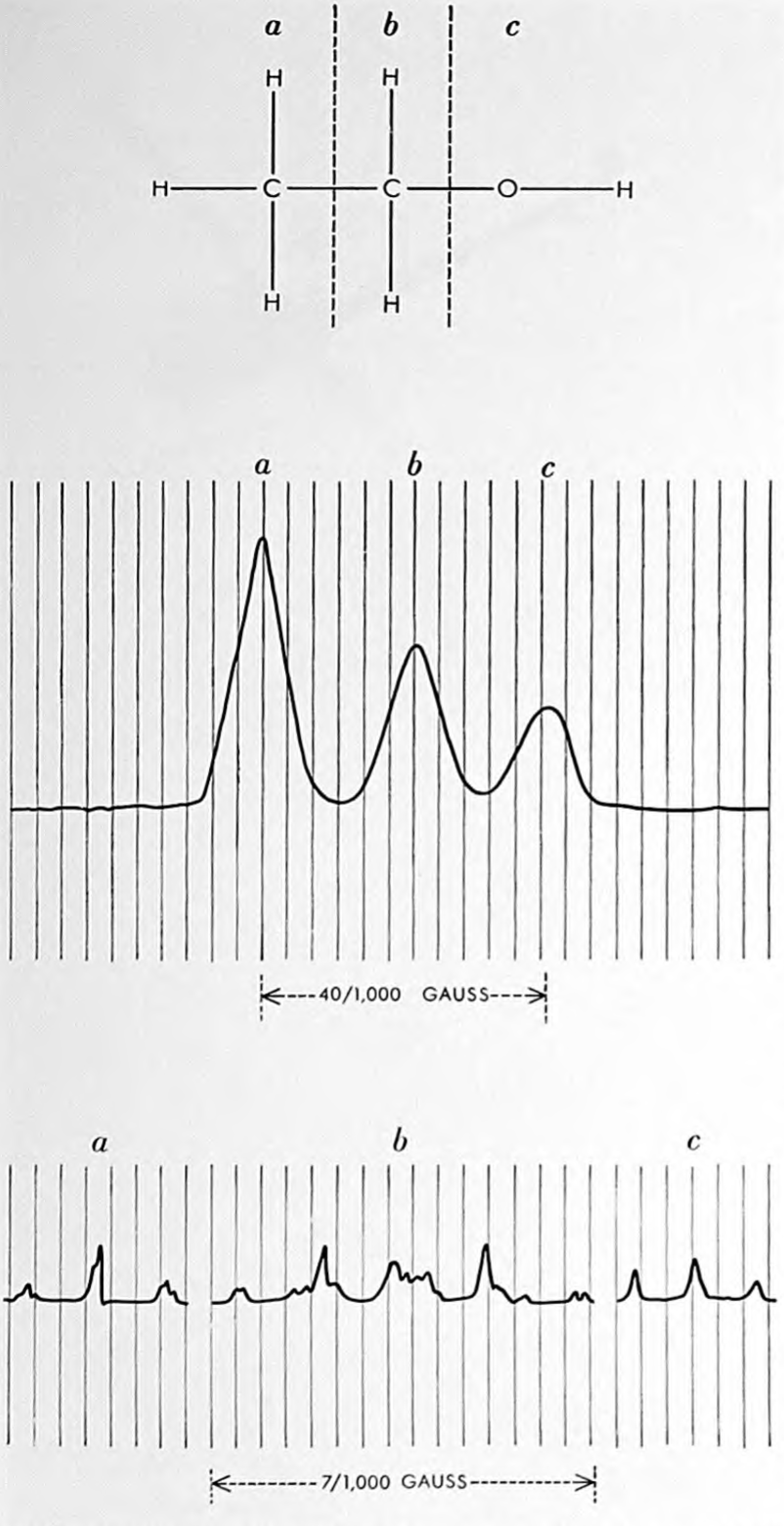
The situation is different in a solid. Because the molecules occupy fixed positions, there are persisting local differences in the magnetic field, and accordingly solids tend to have a broad resonance. The resonance band may be as wide as 20 gauss, as against a sharp line as narrow as one 10,000th of a gauss for some liquids. But here cyclohexane offers an unusual and illuminating case. When it is frozen to the solid state, it still has a sharp resonance down to a temperature of 90 degrees centigrade below its freezing point. This tells us that the molecules in the solid must be in some kind of motion. Evidently they rotate around their positions in the crystal lattice, so that the magnetic field averages out to uniformity.

More detailed study of various substances, solid and liquid, leads to still more interesting developments. In 1947, while working as a graduate student with Purcell at Harvard, I found that the resonance of the hydrogen nuclei in gypsum was split into four distinct lines [*see chart below*]. Clearly these must reflect certain definite variations of the magnetic field within the gypsum crystal. By a theoretical analysis it was possible to translate the information into a picture of how the hydrogen atoms lie in the crystal structure. Here, then, was a new tool for analyzing the structure of crystals—a supplement to probing them with X-rays and neutrons. It is particularly useful for locating light atoms such as hydrogen and lithium, which deflect X-rays only weakly. Nuclear magnetic resonance has now been applied to the study of crystals by scientists in many parts of the world.

From crystals it was a logical step to go on to study the structure of giant molecules, such as rubber, polyethylene and other plastics. One of the problems in analyzing such molecules is to find out how much of their structure is orderly, or crystalline [see "Giant Molecules," a special issue of SCIENTIFIC AMERICAN Offprint 314]. C. W. Wilson, III, one of my graduate students at Washington University in Saint Louis, was able to show that nuclear magnetic reso-



GYPSUM CRYSTAL contains molecules of water, indicated in drawing at left as H-O-H. From the shape of the nuclear magnetic resonance (NMR) spectrum of gypsum (*right*) the positions of the water molecules in the structure of the crystal can be calculated.



ETHYL ALCOHOL, whose chemical formula appears at top, has a three-peaked NMR spectrum (*center*) when examined with a uniform magnetic field. Letters match hydrogen atoms in various sections of the molecule with their corresponding resonance curves. With a still more uniform field the central part of each curve splits further, as shown at bottom.

nance could be used to examine the structure of these huge molecules.

Fingerprints of Molecules

It is in liquids rather than solids, however, that the technique of nuclear magnetic resonance has achieved its most important triumphs of structural analysis. As we have seen, liquids show sharp resonances, because the magnetic field through the material is almost completely uniform. But a few irregularities can be detected when a liquid compound is examined under a very uniform applied field. Under these circumstances it becomes possible to read the resonance spectrum as a "fingerprint" of the structure of the molecule.

For example, James Arnold and Martin Packard at Stanford University, working with an extremely uniform applied field, were able to resolve the resonance of the hydrogen nuclei in the ethyl alcohol molecule (CH_3CH_2OH) into three separate resonances [*see upper chart at left*]. These singled out the three different groups that make up the molecule: CH_3, CH_2 and OH. The explanation is that the hydrogen nuclei in the three groups respond differently to the applied magnetic field because of a shielding effect of the atoms' electrons. The electrons themselves show no magnetism, for in a molecular combination electrons usually are paired off so that each cancels its partner's magnetic field. But an applied magnetic field slightly alters the motions of electrons around their atomic nuclei; the induced motions of the electrons in turn produce weak magnetic fields opposed to the applied field; this "diamagnetism" partly shields the nuclei from the external field. The amount of shielding differs in the different groups of a molecule, and this explains why the hydrogen atoms in the three groups composing ethyl alcohol have different resonances.

With higher resolution (*i.e.*, under a still more uniformly controlled magnetic field) the magnetic spectrum of the ethyl alcohol molecule splits up into an amazing array of separate resonances [*see lower chart at left*]. The spectrum is a fingerprint of the molecule which not only identifies it but also tells much about its structure. Indeed, from such a fingerprint a chemist can sometimes predict the behavior of a molecule.

Chemists in all branches of their discipline are now busily employing nuclear magnetic resonance to unlock the secrets of structure of many kinds of substances, from soap and motor oil to the extremely

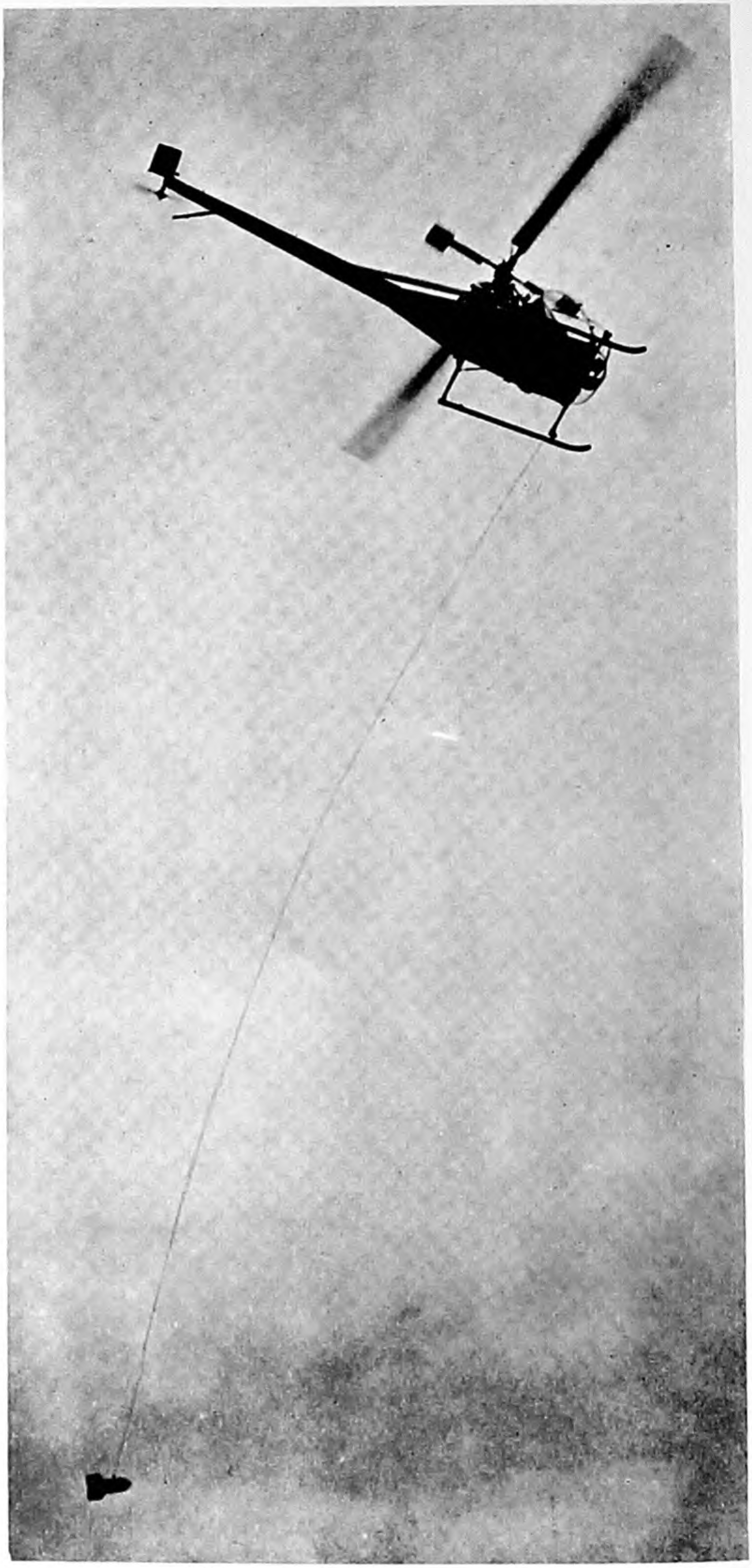
complicated molecules of living matter. One of the virtues of this new analytical tool is that it does not destroy the chemical under analysis. The chemist simply puts the sample in a coil, turns on the radio waves and the magnetic field, records the resonance spectrum and then takes out the sample intact.

Measuring the Earth's Field

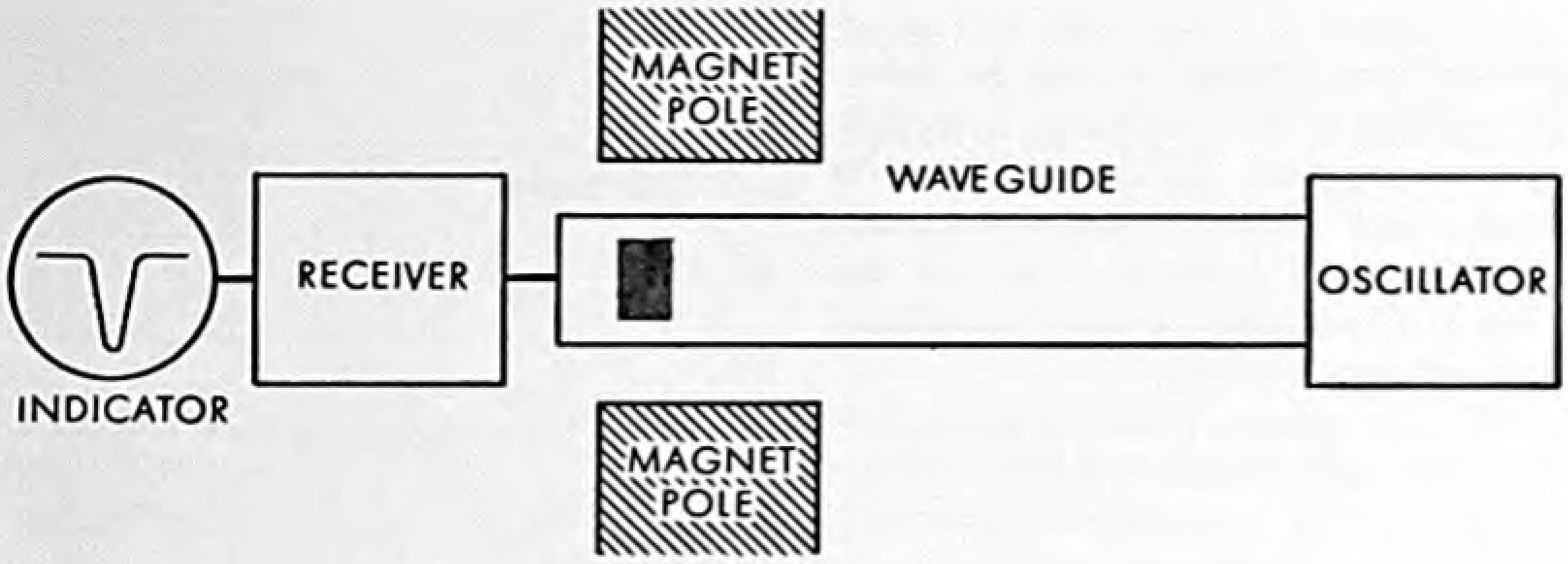
As a final illustration of the versatility of NMR let us look at a very different use—namely, exploring the earth's magnetic field. If very sharp precision in the strength of the applied field is needed to hit the resonance of a liquid, why not reverse the procedure and use resonance for precise measurement of magnetic fields? The difficulty about the earth's field is that it is extremely weak—only about half a gauss. To measure this by the resonance phenomenon, which previously had been studied in the laboratory with fields of thousands of gauss, posed quite a challenge. But Russell Varian and Packard, now with the Varian Associates, have solved the problem with an ingenious device.

Their instrument first lines up the protons in a sample of liquid with a moderately weak magnetic field. The protons are all oriented in one direction so that the whole sample, in effect, is a weak magnet. Then the polarizing field is suddenly switched off; the magnetic rug is pulled out from under the group of protons, so to speak. The nuclear magnets, which have been lined up in a direction *not* parallel to the earth's field, now begin to precess around the axis of this field. Their alignment rapidly breaks down, but in liquid benzene it lasts up to 20 seconds. This is long enough to measure the strength of the earth's magnetic field to an accuracy approaching one part in 10 million. The measurement is made simply by tuning in to the precession rate of the group of protons: since the earth's field produces the precession, the rate is a measure of the strength of the field.

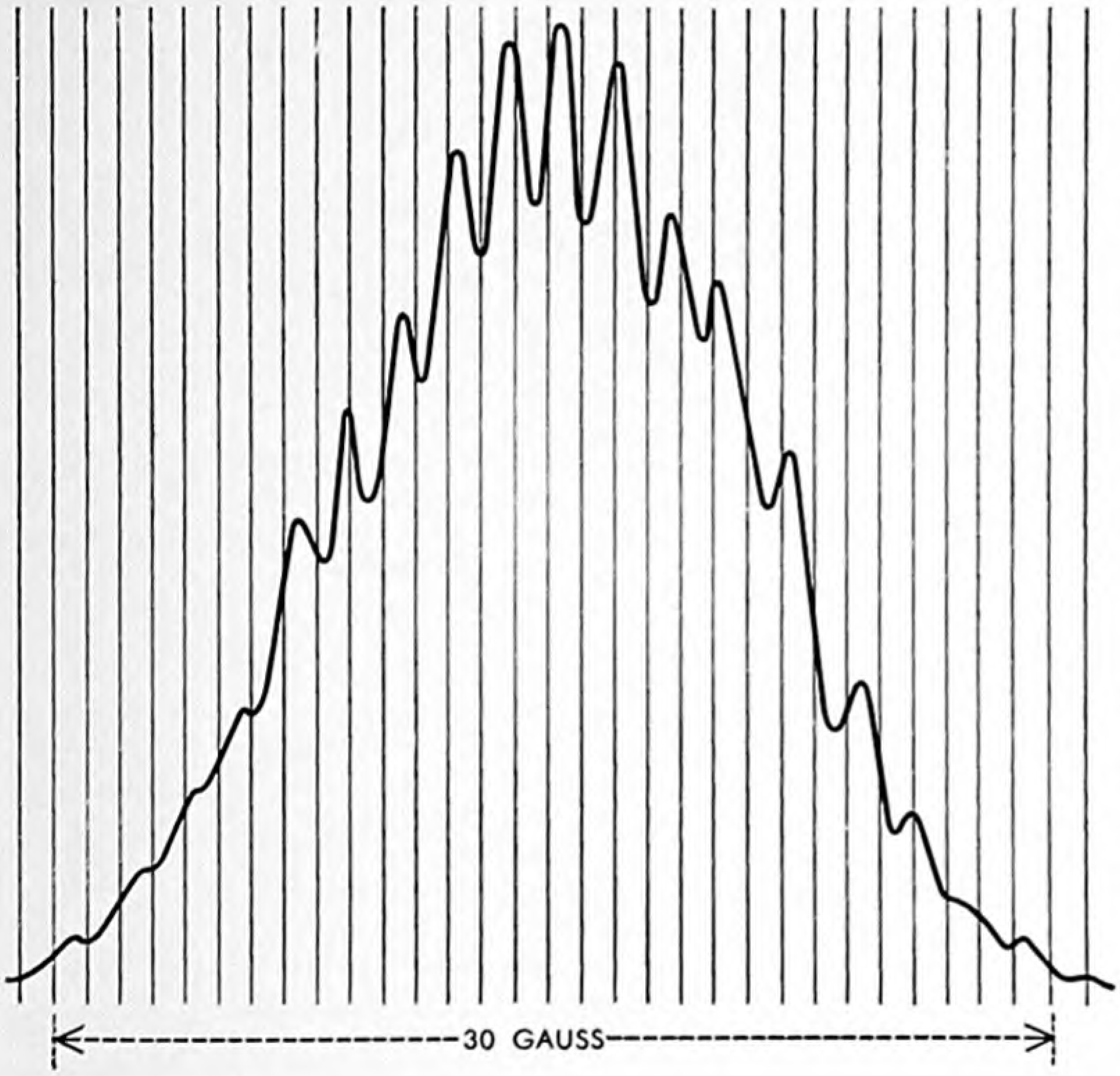
Varian and Packard named their instrument the proton precessional magnetometer. Obviously this amazingly sensitive device could serve to measure variations in the earth's magnetic field. It has already been put to use in prospecting for mineral deposits, from the air and on the ground [*see photograph at right*]. The instrument has also been shot in rockets to measure the strength of the earth's field at various heights above the surface. A magnetometer of this type is scheduled to go up in

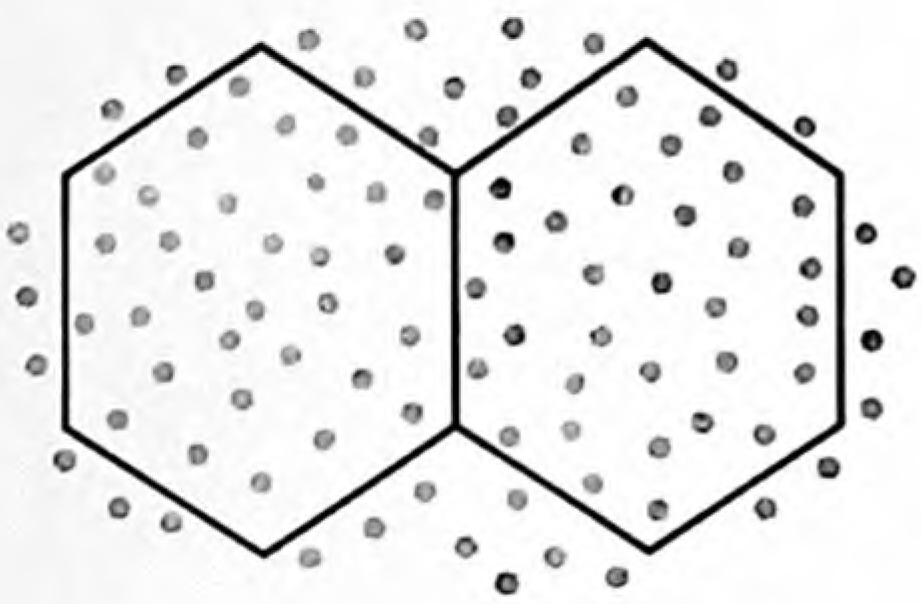


MAGNETOMETER based on nuclear magnetic-resonance principle measures small variations in the earth's field. Here it is towed by a helicopter in a mineral-prospecting survey.



ELECTRON-RESONANCE APPARATUS operates on the same principle as Purcell's circuit seen on page 273. A microwave oscillator sends three-centimeter radio waves through the waveguide. The sample (*gray rectangle*) is mounted on the inside of the guide. Resonance is indicated by a decrease in the power delivered by the waveguide to the receiver.





HYPERFINE STRUCTURE (jagged splitting) of an electron-paramagnetic-resonance (EPR) spectrum indicates various positions of the unpaired electron in a free-radical molecule. This curve comes from naphthalene radicals. Gray dots in the structural outline of the radical at bottom indicate that the free electron wanders over many parts of the molecule.

one of the U. S. artificial satellites soon. Surely Purcell and Bloch could hardly have foreseen that their exploration of the magnetism of the proton would lead to such developments.

The Resonance of the Electron

Let us turn now to the magnetism of the electron. The magnetic resonance of this particle was discovered by a Soviet physicist, E. K. Zavoisky, in 1944, before Purcell and Bloch tuned in on the proton. But curiously enough the Russians apparently were not as quick to exploit the discovery as physicists in the U. S., Great Britain and the Netherlands. At all events, the electron's resonance, called electron paramagnetic resonance (EPR), has now been forged into a tool as important in its area of usefulness as nuclear magnetic resonance.

As we have seen, the electron, because of its smaller mass and faster spin, is a much stronger magnet than a proton. As a result, a given magnetic field makes it precess far more rapidly. Its precession rate in the standard laboratory magnetic field is in the range of the frequency of radio microwaves—that is, about 10,000 megacycles per second, or a wavelength of about three centimeters.

When microwaves travel down a rectangular waveguide (the tube used to conduct such waves), they produce a rotating magnetic field at any fixed point. This field can serve to flip over the electron magnets in matter, just as a rotating field in a coil flips protons. The experimenter may place the sample of material on a side wall of the waveguide, turn on the radio waves and apply an external magnetic field to make the electrons precess. When the precession rate reaches the resonance value and the electrons flip, they extract energy from the radio waves, and the reading on a receiver at the end of the tube dips accordingly.

Now this technique obviously can tell us nothing about substances in which the electrons are all paired, *i.e.*, where the electrons' magnetism is neutralized. But it has proved very helpful indeed in studying material with unpaired electrons. Electron resonance was first applied to investigate crystals containing elements with unfilled electron shells (therefore unpaired electrons)—elements such as manganese and iron. Much has been learned about substances of this kind, particularly about those used for "magnetic" cooling of matter to very low temperatures.

But more exciting has been the discovery that electron resonance can be used to investigate free radicals, the

transitory molecular fragments that play a crucial role in many chemical processes, including the chemical activities of living cells [see "Free Radicals," by Paul D. Bartlett; SCIENTIFIC AMERICAN, December, 1953]. Free radicals, of course, have unpaired electrons, and with the electron-resonance technique it is sometimes possible to detect these fleeting substances and to learn something about their structure and behavior.

At first thought one might suppose that the resonance spectrum of the unpaired electron in a free radical should always be the same—one free radical indistinguishable from another. But this is not the case. The electron is affected by the magnetic field of the nuclei in whose neighborhood it happens to be, and as the free electron wanders about in the molecular fragment, it is subjected to varying magnetic fields. As a result its resonance may be split into a "hyperfine structure" [*see chart on page 277*]. From the splitting we may learn where the electron spends its time and at what rate the free radical is likely to enter into chemical reactions.

Such studies are not limited to natural free radicals. With high-energy particles from an accelerator it is possible to break a molecule into fragments, and the fragments can sometimes be frozen in their tracks, so to speak, by keeping the sample at a very low temperature. We can then see what has happened to the molecule by examining the electron resonances of the fragments. This kind of investigation could be useful for studying the chemical effects of radiation on certain plastics; irradiation is known to strengthen some plastics by causing them to form new chemical bonds. Magnetic resonance also looks promising as a tool for investigating the free radicals that catalyze the synthesis of high polymers such as rubber or polyethylene [see "How Giant Molecules Are Made," by Giulio Natta; SCIENTIFIC AMERICAN, September, 1957].

Resonance in Living Cells

To illustrate the interest of biologists in magnetic resonance, I shall close this account with an episode that began in Saint Louis one evening in 1951. After attending a chamber music concert, a few members of the Washington University faculty repaired to someone's home for coffee, and the conversation turned (probably to the wives' chagrin) to shop talk. Barry Commoner of the botany department fell to discussing the theory that free radicals play an important role in the processes of oxidation and reduction in living cells, and he remarked how difficult it was to detect free radicals in living systems. I suggested that electron resonance might be helpful, and offered to help Commoner and his group learn how to use the method. Thus began a most interesting series of experiments.

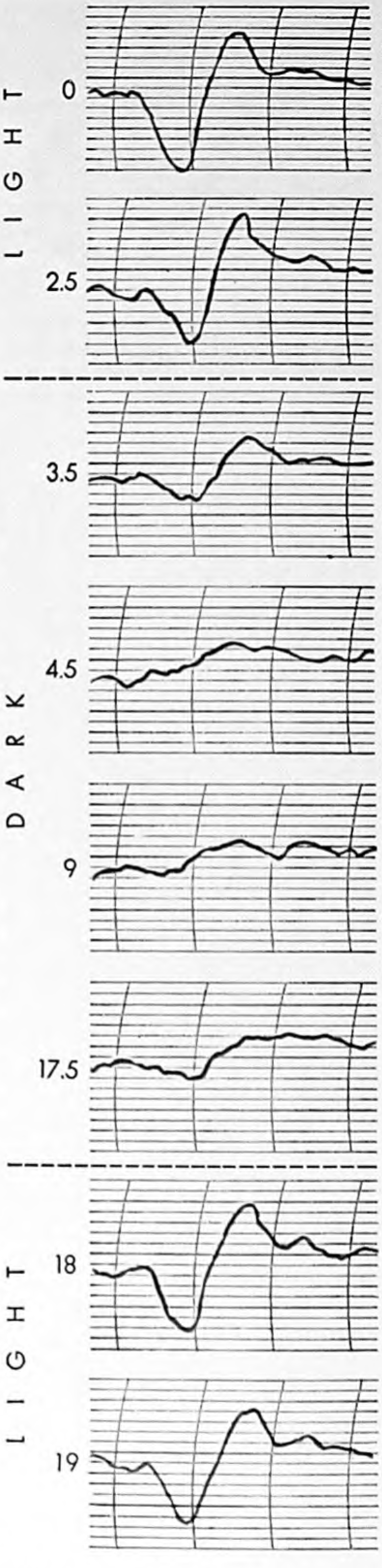
The first results were disappointing. Since the experiments involved cultures of cells (yeast and other organisms), there was moisture in the microwave apparatus, and moisture absorbs microwaves. This absorption of course masked the dip in radio energy produced by magnetic resonance. Even so, the experimenters finally detected a weak resonance from yeast and found that its intensity varied with the rate at which oxygen was consumed by the yeast cells. Then a resonance was discovered in green leaves ground up and quickly freeze-dried to eliminate the absorption by moisture.

To study the active process of photosynthesis in its necessary moist environment, however, called for a much more sensitive apparatus. Jonathan Townsend of the physics department succeeded in devising one, and it became possible to work with living cells. The cells selected contained chloroplasts—the small, green, chlorophyll-packed bodies which are thought to carry out most if not all the steps of photosynthesis.

The cells were placed in the sensitive magnetic resonance apparatus and were simultaneously irradiated with white light from a 50-candle-power lamp. The chloroplasts immediately showed a dramatic resonance spectrum [*see charts at right*]. When the light was turned off, the resonance soon weakened or disappeared entirely. Next the cells were exposed to light of various specific wavelengths; it turned out that the resonance appeared at the very same range of wavelengths of light that produces photosynthesis.

Commoner and Townsend have gone on to further experiments which not only have linked free-radical activity firmly to photosynthesis but have also indicated that free radicals are involved in the metabolism of cancer cells. And biological investigators in many other laboratories have begun to adopt electronic resonance as a tool in their researches.

That basic discoveries in science invariably bear fruits which cannot be foretold is an old story to scientists. Even so it has been a great thrill to see what has grown out of the work of Purcell, Bloch and Zavoisky, who were seeking only to get a better understanding of the magnetism of the particles in the atom.



CHLOROPHYLL SAMPLE **gives sharp EPR spectrum when irradiated with light. This proves that free radicals are involved in photosynthesis. Numbers at left indicate minutes from the start of the experiment.**

The Author

GEORGE E. PAKE was Edward Purcell's graduate student at Harvard University when Purcell was just beginning his Nobel-prize-winning work on magnetic resonance. Pake, who is now professor of physics at Stanford University, has been investigating magnetic resonance ever since. Pake grew up in Kent, Ohio, where his father taught English at Kent State University. As a student in the University's "laboratory" high school for teacher training, he had access to the University library, and there his interest soon turned from baseball and model airplanes to physical science. With the help of Kent State's faculty, Pake received a head start in physics which served him well when, as a Westinghouse Scholarship student at the Carnegie Institute of Technology, he received his B.S. and M.S. degrees just 32 months after matriculating. "My undergraduate career whizzed by in a kind of blur," Pake reports. "I often think I would never have survived the concentrated dose of studies had it not been for my music and the opportunity to play French horn in the Carnegie Tech orchestra." Pake worked at Westinghouse Research Laboratories for a year, then went to Harvard, where he received his Ph.D. in 1948. Four years later, at the age of 28, he became chairman of the physics department at Washington University in Saint Louis. He has been at Stanford since 1956.

Bibliography

NUCLEAR AND ELECTRONIC SPIN MAGNETIC RESONANCE. John E. Wertz in *Chemical Reviews*, Vol. 55, No. 5, pages 829-955; October, 1955.

NUCLEAR MAGNETIC RESONANCE. E. R. Andrew. Cambridge University Press, 1955.

NUCLEAR MAGNETIC RESONANCE. G. E. Pake in *Solid State Physics*, Vol. 2, pages 1-91; 1956.

NUCLEAR MAGNETISM IN RELATION TO PROBLEMS OF THE LIQUID AND SOLID STATES. E. M. Purcell in *Science*, Vol. 107, No. 2,783, pages 433-440; April 30, 1948.

SPECTROSCOPY AT RADIO AND MICROWAVE FREQUENCIES. D. J. E. Ingram. Butterworths Scientific Publications, 1955.

SCIENTIFIC AMERICAN October 1954, Vol. 191, No. 4, pp. 40-44 OFFPRINT 234

THE LINEAR ACCELERATOR

by Wolfgang Panofsky

The early idea of accelerating particles in straight lines instead of circles has made a strong comeback. Linear machines generate electron or proton beams of respectable velocity and sharp focus.

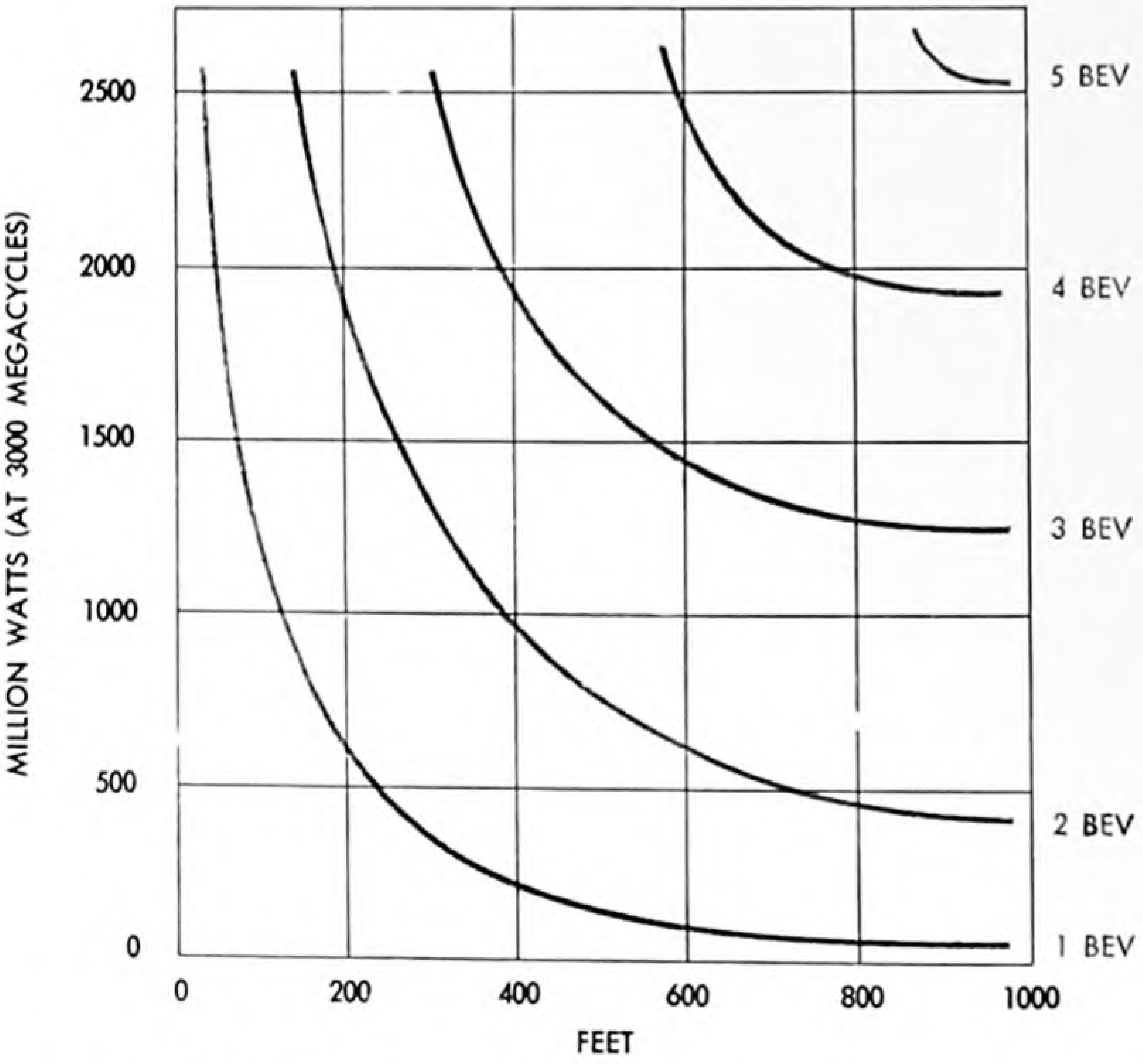
A quarter of a century ago physicists started a "race" toward more and more energetic machines for the acceleration of atomic particles. The evolution of particle accelerators has now reached a stage where specialization of the machines for the specific purposes for which they are to be used has become as important as brute energy. The linear accelerator, which is the subject of this article, falls in the category of a specialized tool.

The idea of linear acceleration actually is older than the cyclotron or the present more potent successors of the cyclotron. Indeed, the linear accelerator was once believed to be the most promising candidate to win the energy race [see "The Bevatron," by Lloyd Smith; SCIENTIFIC AMERICAN, February, 1951]. It is still the "favorite" to attain the highest energy in the acceleration of electrons, but at the present time it is not a leading contender for the maximum acceleration of protons; in that field circular machines have taken the lead as a result of improvements in their design and certain practical difficulties of the linear accelerator. The main attractions of the linear machine now, however, are not its peak energy possibilities but other qualities which we shall examine.

The linear accelerator, as its name implies, accelerates particles in a straight line; this distinguishes it from all other present accelerators, which drive the particles around a circular track. Like the other machines, the linear one speeds up the particles by means of a series of boosting kicks, or, in another version, by means of a continuously accelerating electrical wave.

The first of these two forms is known as the "drift-tube" accelerator. The particles travel through a long series of pipes, and at each gap between one pipe and the next they are given a boost by a voltage applied in the right direction at precisely the right time. The energy and timing of the alternating voltage are provided by radio waves from powerful transmitters.

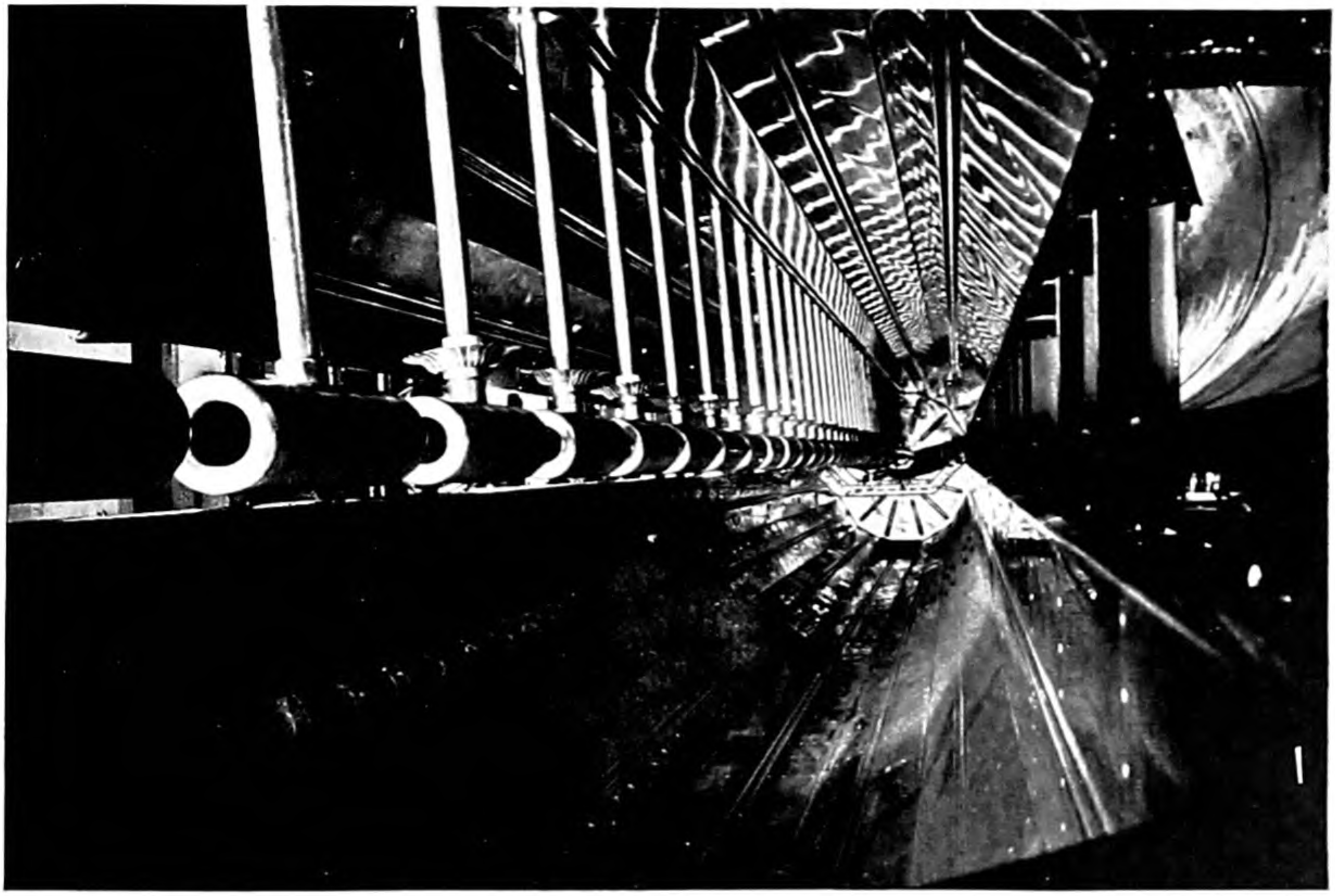
The second form of the linear machine is called a "traveling-wave" accelerator. Here the long pipe is continuous. An electric field generates an electromagnetic wave in it, and the wave, controlled in speed to stay with the particle, carries it along to steadily higher velocities. The effect is analogous to a water wave carrying a surfboard rider.



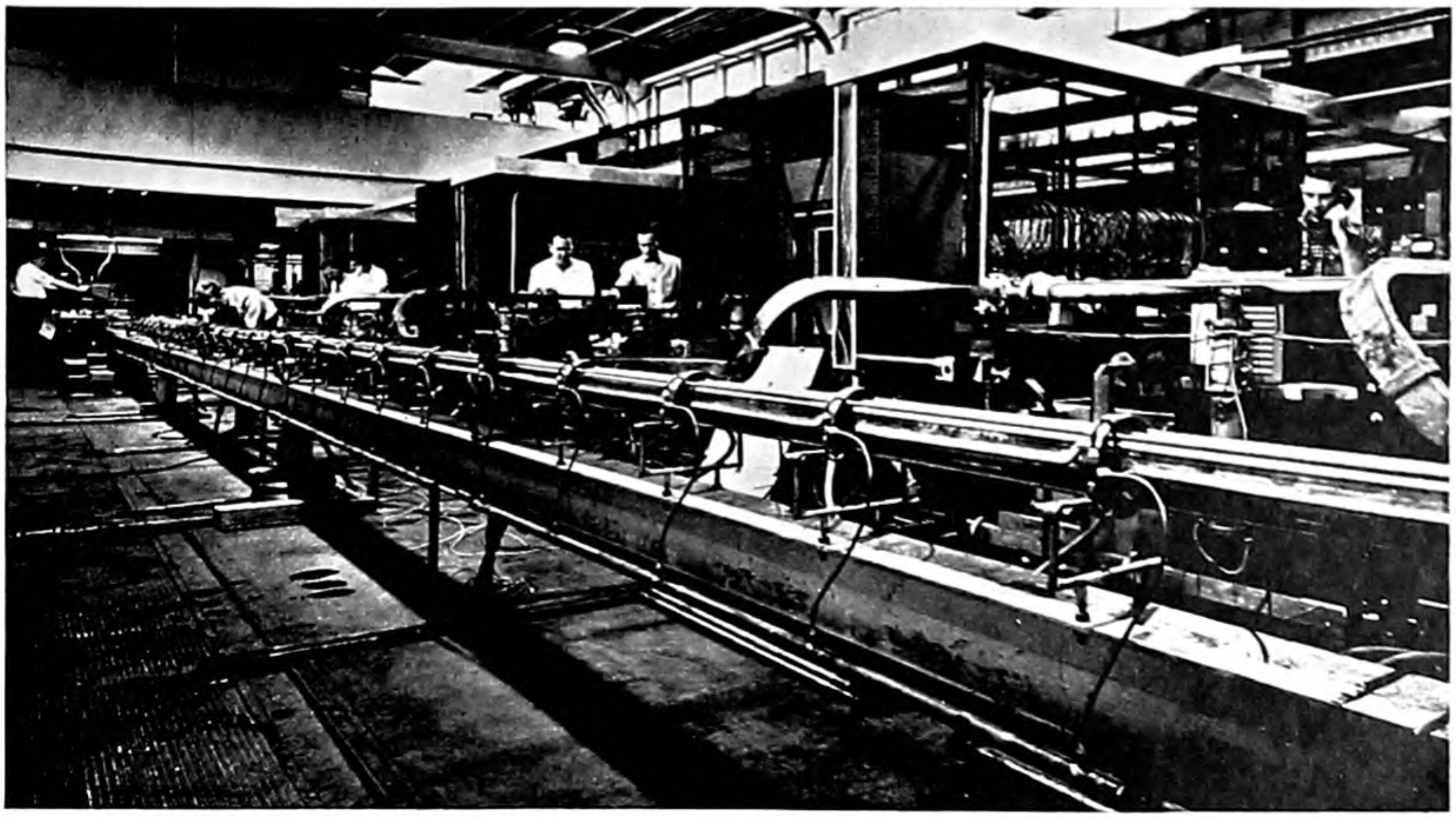
ENERGIES attainable from the linear electron accelerator operating at 3,000 megacycles are plotted in billions of electron volts. Length can be traded for power on equal terms.

The idea of a linear accelerator was suggested as early as 1924 by Gustaf Ising in Sweden. Ernest O. Lawrence and David H. Sloan at the University of California were the first to build one. They used radio-frequency power. Because of the limitations of radio generators of that time they could not accelerate particles to very high velocities.

The late W. W. Hansen of Stanford University became interested in such a machine and realized that success would depend on obtaining large amounts of



BERKELEY ACCELERATOR of the internal drift-tube type is shown with the resonator cavity open. Supports locate the drift tubes precisely in the proton path. Protons spend one whole cycle traveling through each tube. They develop energy up to 32 Mev.



STANFORD ACCELERATOR of the traveling-wave type attains energies of 600 Mev. Each feed point every 10 feet delivers 20 million watts of power from a klystron generator. The present total length is 220 feet. The usual shielding is absent in this picture.

radio-frequency power and on generating high voltages at the minimum expense of power. The latter is necessary because a linear accelerator must apply power at many different gaps instead of at only one, as in the cyclotron. Hansen therefore looked for a more efficient means of generating high radio-frequency voltages, and he hit upon the cavity resonator, which was translated into the klystron tube [see "The Klystron," by Edward L. Ginzton; SCIENTIFIC AMERICAN, March].

Hansen had set out to find the means of realizing a practical linear machine for accelerating electrons; his work actually led to the development of powerful radio-frequency generators and to microwave radar and radio. After the war these developments in turn led, by feedback, to the realization of linear accelerators not only for electrons but also for protons and heavier particles. Luis W. Alvarez of the University of California and others reopened the possibilities of this type of machine.

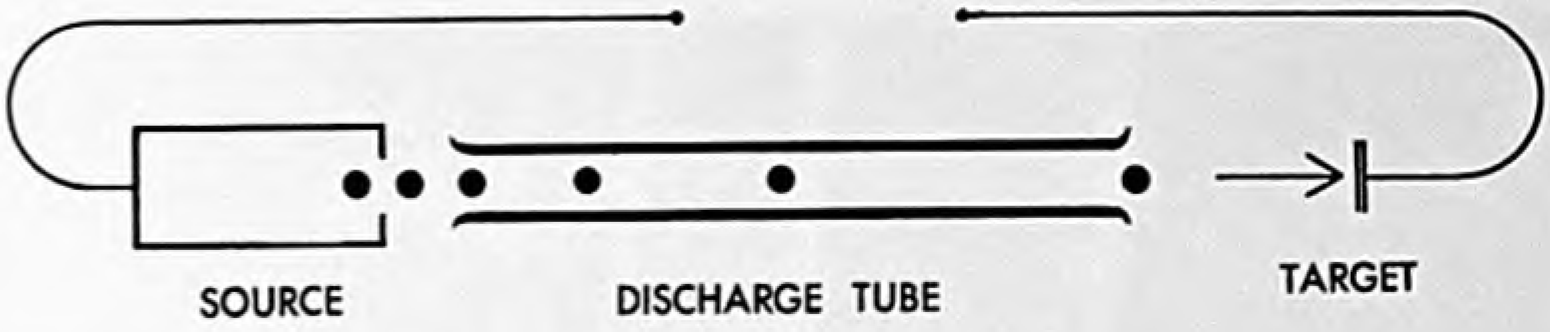
Some experiments in physical research are done best by a circular accelerator, others by a linear machine. In gauging the suitability of a machine for a given task, the nuclear physicist must consider a number of factors: the kind of particle to be accelerated, the maximum velocity to which the machine can accelerate them, how large a current (*i.e.*, the number of accelerated particles per second) it can deliver, how well it can focus the beam on the target, how sharp an energy spectrum it can provide, and so on. And of course there is the factor of cost, not only the initial cost of building the machine but also the costs of operating and maintaining it.

Let us see how the linear accelerator compares with the circular types in these respects. If electrons are the particles to be accelerated, the linear accelerator can reach higher energies and deliver larger currents than any other machine. It can also train a beam on the target more easily. On the debit side of the ledger are the facts that its energy spectrum is generally less sharp and that its power requirements restrict its operating time to shorter pulses and create engineering problems which have not yet been solved as satisfactorily as those of the circular machines. As a consequence a linear accelerator at present is less reliable and costs more to maintain. When it comes to accelerating protons, the linear accelerator in principle can reach any desired level of energy; however, the once-held idea that linear machines in the high-energy range would be less costly to build has proved wrong, partly because innovations such as the strong-focusing principle have reduced the cost of the circular types [see "A 100-Billion Volt Accelerator," by Ernest D. Courant; SCIENTIFIC AMERICAN, May, 1953].

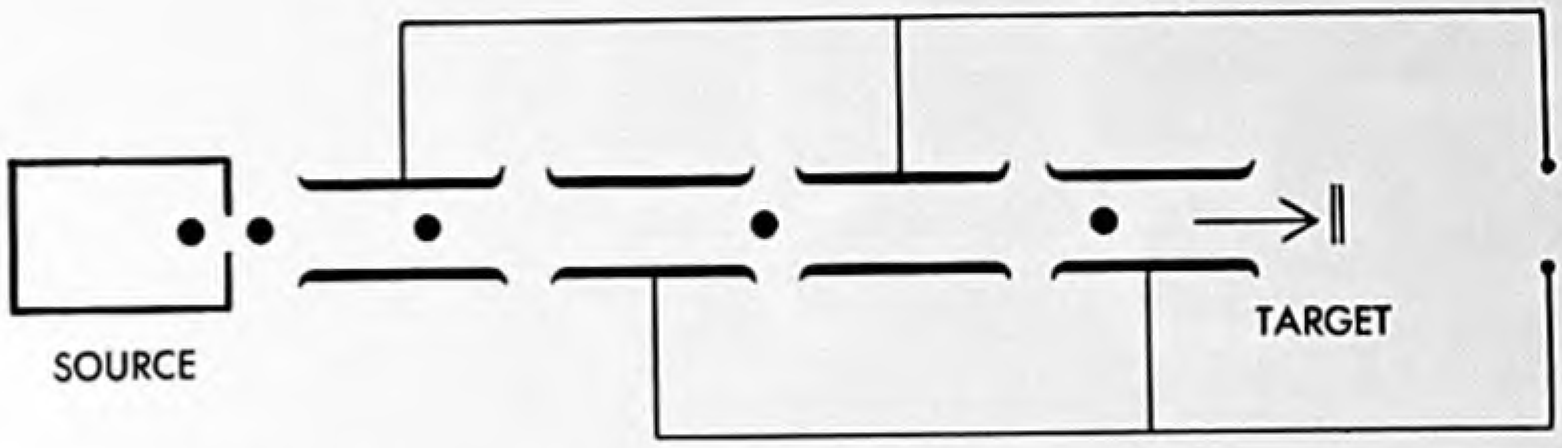
Nonetheless for certain uses the linear accelerator offers such marked advantages—particularly the ease with which it delivers a well-collimated beam of accurately regulated energy on the target—that it is sometimes the best choice. For example, the British Atomic Energy Research Establishment, after considering whether to build a three-billion-electron-volt circular accelerator or a 600-million-electron-volt linear accelerator for protons, recently decided in favor of the linear machine.

The illustrations accompanying this article show some of the principles and components of the various types of linear accelerators. In the first machine, as suggested by Ising and worked out by Sloan and Lawrence, an electric generator discharged positively charged particles into a series of drift tubes [*see second drawing above*]. As a positively charged particle leaves the source, tube 1 is negative, tube 2 is positive, tube 3 is negative, and so forth. By the time the particle reaches the end of tube 1, the voltage has reversed, so that tube 1 is positive and tube 2 negative. Thus the particle continues to be accelerated. This process continues through the length of the machine. The lengths of the tubes have to be tailored so that the particle spends exactly one half cycle in each drift-tube space. This essentially is the machine which, before the war, accelerated mercury ions to two million electron volts (2 Mev).

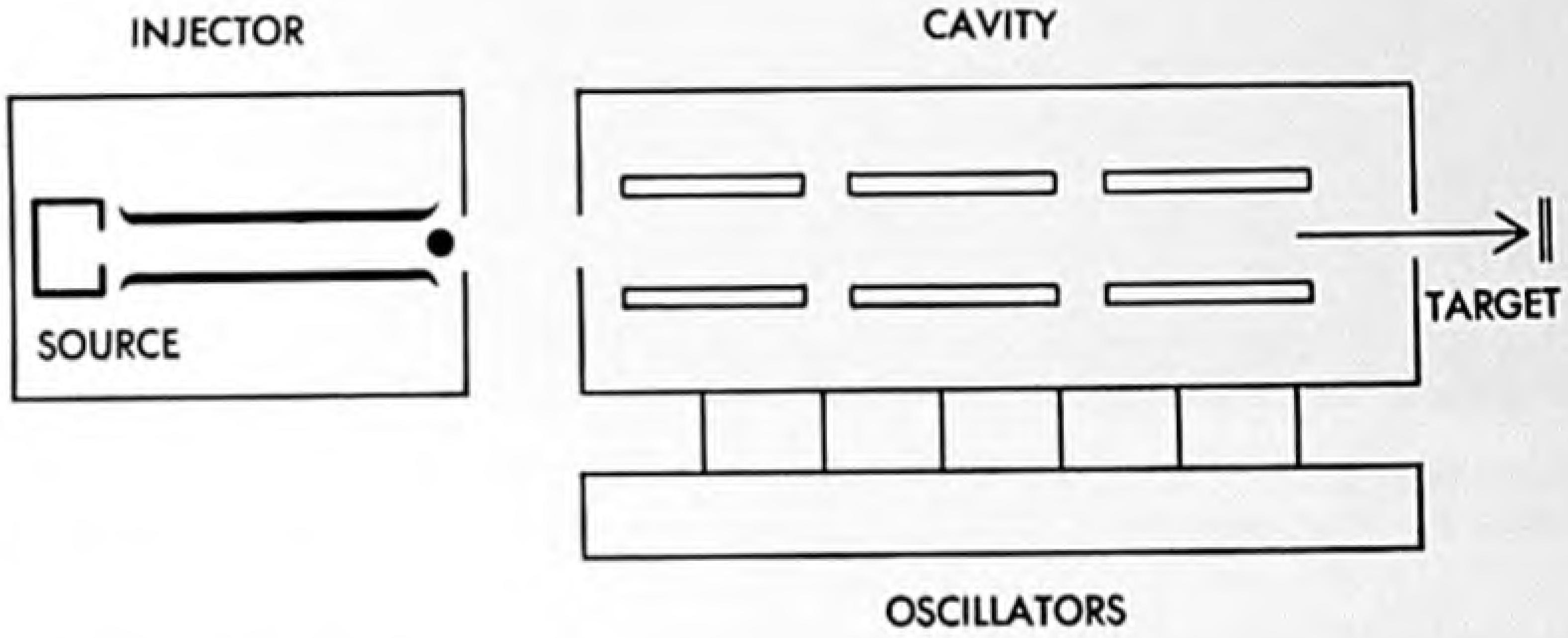
The limiting factor on a structure of this kind is the power source. Particles of interest to nuclear physicists travel with a large fraction of the speed of light;



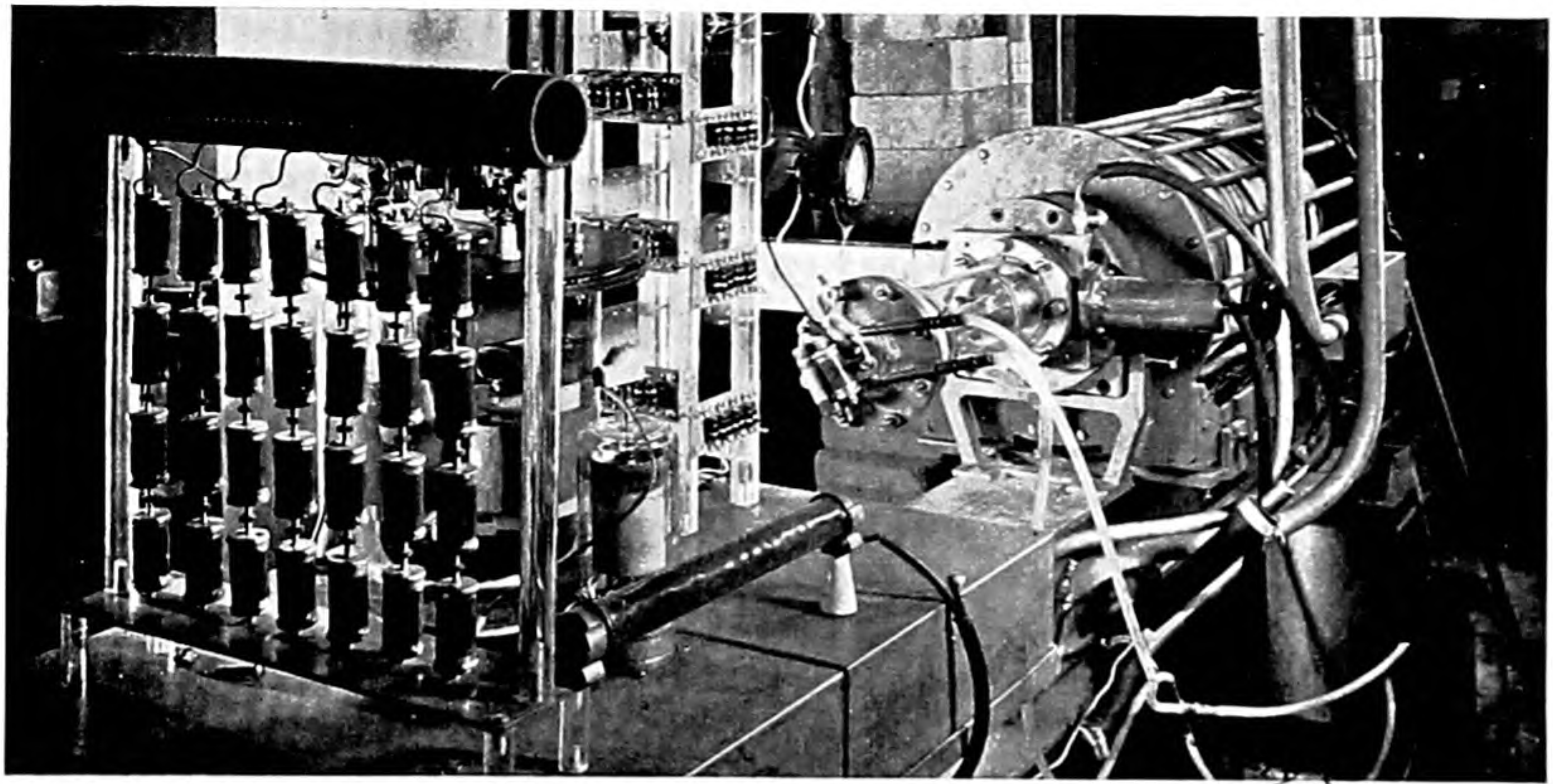
STATIC ACCELERATOR is simplest design. The particles are generated by an electron or ion source. A high static voltage accelerates the particles down a discharge tube.



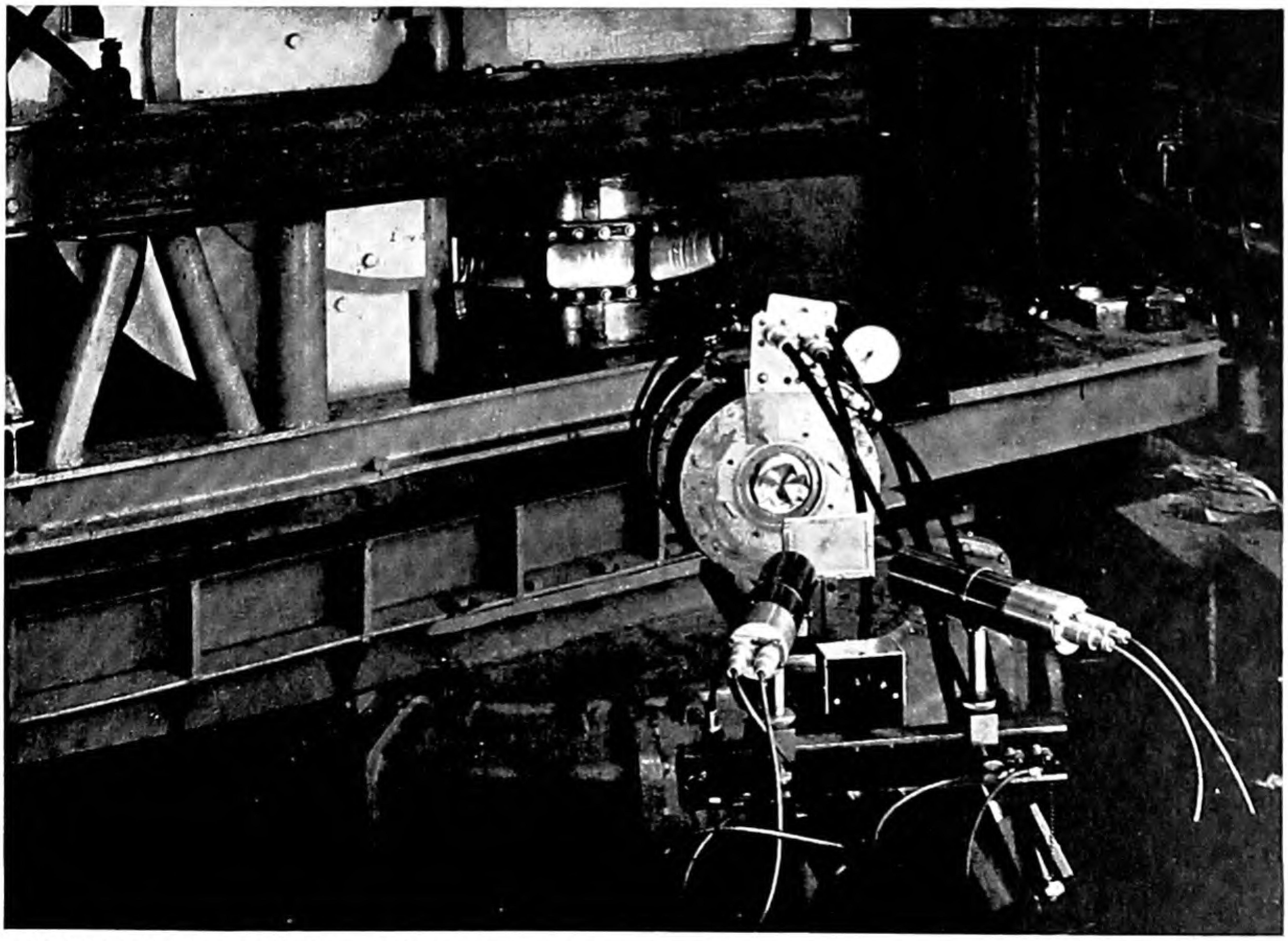
DRIFT-TUBE ACCELERATOR of an externally fed type accelerates particles down a series of tubes by alternating voltage so that successive tubes are of right charge as particle passes.



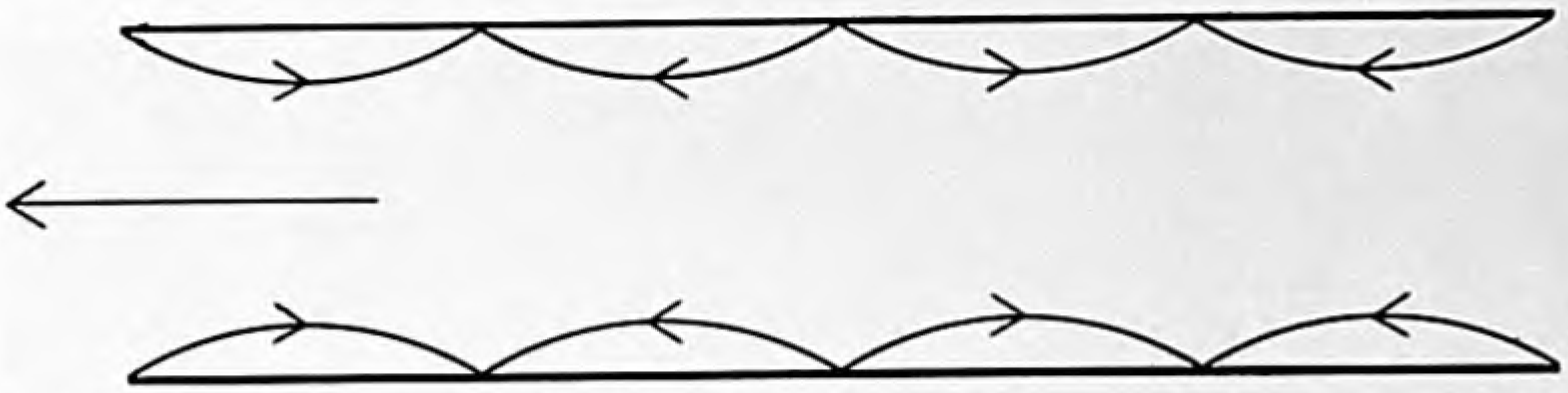
CAVITY ACCELERATOR employs drift tubes, but radio-frequency oscillators apply alternating voltage along cavity. Timing is such that particles meet accelerating field in each gap.



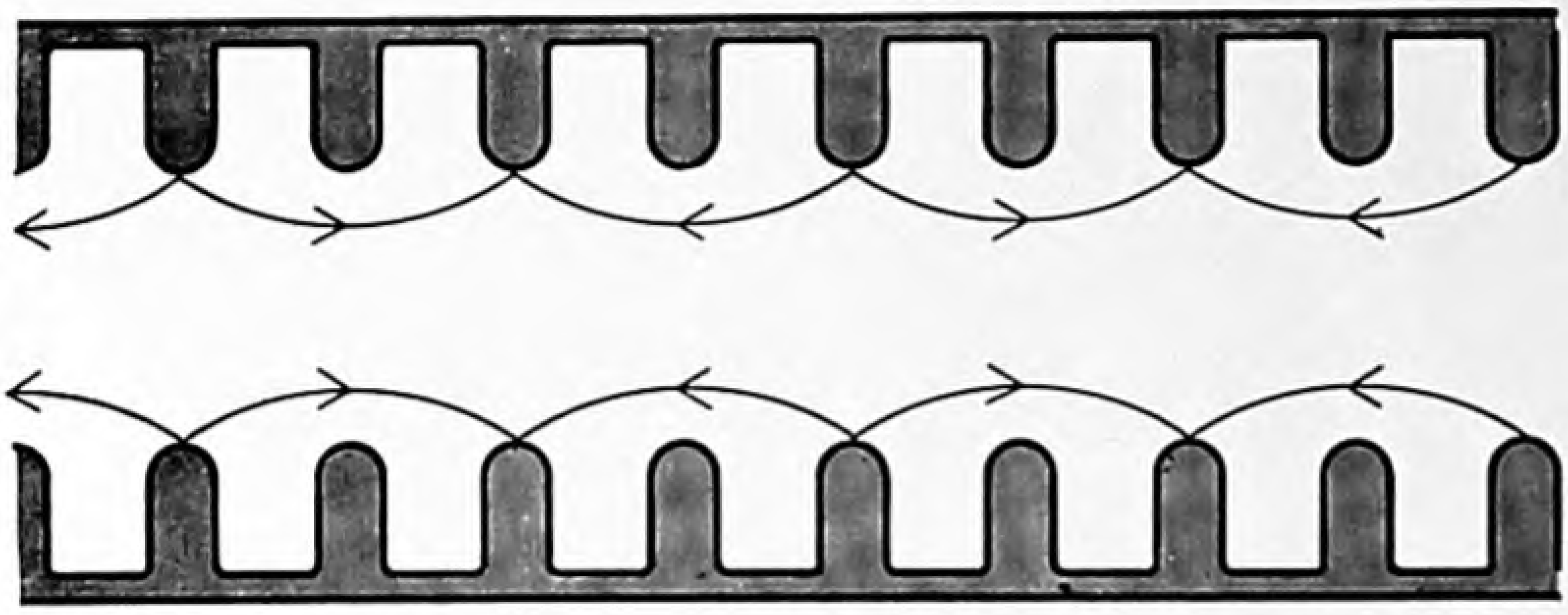
SOURCE of electrons for the Stanford 600-Mev linear accelerator is a hot filament housed at the beginning of the traveling-wave tube (*right*). As the electrons are evaporated off the filament, they are given an initial velocity of 80 kilovolts by a pulser (*left*).



TARGET installation at Stanford has the beam enter a scattering chamber (*bottom*) where it is scattered at various angles. Electrons are drawn by the semicircular magnet to a shielded counter (*above*). The latter can be turned on its rotating gun mount.



UNIFORM WAVE GUIDE is diagrammed showing the generated electromagnetic field with components parallel to the tube axis. In this tube the field travels faster than light.



LOADED WAVE GUIDE, of corrugated type, works by placing load on the wave. This slows the wave to the speed of the particles, which, for electrons, approaches the speed of light.

hence the voltage on the electrodes has to be alternated very rapidly. This requires a large current flow through the wires, which in turn results in large power losses. And the high-frequency power is expensive to generate.

Hansen's cavity resonator reduced the power losses by distributing the currents over the large conducting surfaces of the cavity instead of concentrating them in wires where much of the energy was lost as heat. The lower drawing of the three on page 282, illustrating the operation of the proton linear accelerator at Berkeley, shows how a cavity-controlled system works. Power is fed to the cavity from banks of high-powered radio transmitters. The cavity transforms this radio energy into an alternating electric field which makes currents rush back and forth across the gaps between the ends of the drift tubes in the series. At any instant the direction of the accelerating force is the same in all the gaps. If the structure is designed so that a group of particles takes the time of one full cycle of the alternating voltage to travel through each drift tube, then the particles will always find themselves in an accelerating field when they cross a gap.

Although a structure of this design is much more efficient than the first one, it still requires very high power. In order to hold down the average power consumption to a reasonable level, the power must be applied in very short pulses. The 32-Mev proton linear accelerator at Berkeley requires 2.2 million watts of radio-frequency power (near the ultrahigh-frequency television band). It is powered by eight individual transmitters. This machine has been in continuous operation for several years and has been a very valuable tool in physics. A 70-Mev machine of similar design is nearly completed at the University of Minnesota, and similar principles will be used in the 600-Mev machine at Harwell, England.

The traveling-wave type of linear accelerator is particularly applicable to electrons. The wave travels in a cylindrical pipe with conducting walls, known as a waveguide. Of the various kinds of traveling waves used, a particularly interesting type is the so-called TM wave, in which the electric field always has a component parallel to the axis of the cylinder [*see upper drawing above*]. Imagine now that this whole field structure travels along the axis. If any particles happened to be moving with the same speed, they would be accelerated. For theoretical reasons which we need not go into here, electromagnetic waves cannot be slowed down to the maximum speed at which particles can travel in a simple, uniform pipe. To slow them down we must use some sort of interruption; the most popular method is to introduce disks of conducting material at regular intervals along the pipe, forming what is known as a "corrugated" waveguide [*lower drawing above*]. The disk spacing, disk thickness, disk aperture and pipe radius all have to be chosen to give minimum power loss at the needed wave velocity. The structure has to be fabricated with high precision (sometimes within a tolerance of .0002 of an inch), for the wave speed is very sensitive to small variations in the dimensions. The general dimensions of the tube also depend on the kind of particles to be accelerated; a proton machine must be much larger than one for electrons.

The corrugated-guide linear accelerator has been advanced furthest at Stanford University. Two machines are operating: one, 10 feet in length, attains 38 Mev; the other has reached 600 Mev and is expected to go higher. In the latter machine the accelerating tube itself is dwarfed by its auxiliary equipment, its shielding and the facilities for performing experiments on target material [*see photograph on page 281*]. This emphasizes the point that the economics of high-energy machine physics often depends only in small part on the design of the accelerating device.

The 600-Mev machine pictured is designed for a total power input of 400 million watts (roughly 20 million watts for each 10-foot section). Such power, at a frequency of 3,000 megacycles, is not available from commercial sources at this time. It is generated at Stanford by a set of klystron amplifiers driven from a common high-frequency source.

The British have been notably successful in commercial development of this type of accelerator for uses that do not require such high power. They have built research machines yielding four Mev and 15 Mev, and several firms are producing machines of four Mev energy for cancer therapy and radiography.

Essentially there are two ways to go to high energy in a linear accelerator: build a bigger machine or put in more power. The choice between these is fundamentally an economic question: whether the extra construction or the extra power will be more costly. It appears in retrospect that the Stanford machine might have been slightly more economical if we had made the tube longer and needed less power. The designer of a linear accelerator has to estimate how much a megawatt of power is going to cost him and what the cost per foot of machine would be; an optimum for a given output energy can then be calculated.

This article has given only a general outline of technical features of linear accelerators and has not attempted to discuss their applications in detail. It is clear, however, that these machines should be valuable tools in the struggle to understand high-energy phenomena.

The Author

WOLFGANG PANOFSKY is professor of physics at Stanford University. He was born in Berlin, the son of art historian Erwin Panofsky, who later came to the Institute for Advanced Study at Princeton. After graduating from Princeton University in 1938, young Panofsky went to the California Institute of Technology, where he took a Ph.D. in physics and stayed to direct a project for the Office of Scientific Research and Development. From 1945 to 1951 he taught physics at the University of California, and since then he has been at Stanford. He is director of Stanford's High-Energy Physics Laboratory.

Bibliography

RECENT PROGRESS IN ACCELERATORS. E. L. Chu and L. I. Schiff in *Annual Review of Nuclear Science*, Vol. 2, pages 79-92; 1953.

LINEAR ACCELERATORS. D. W. Fry and W. Walkinshaw in *Reports on Progress in Physics*, Vol. 12, pages 102-132; 1948-1949.

A LINEAR ELECTRON ACCELERATOR. E. L. Ginzton, W. W. Hansen and W. R. Kennedy in *The Review of Scientific Instruments*, Vol. 19, No. 2, pages 89-108; February, 1948.

SCIENTIFIC AMERICAN October 1954, Vol. 191, No. 4, pp. 40-44 OFFPRINT 235

MODELS OF THE NUCLEUS

by R. E. Peierls

How do physicists presently visualize it? Curiously, different approaches to the nucleus suggest different pictures, notably the liquid-drop model, the shell model and the optical model.

Ever since 1930, when the discovery of the neutron made it plain that the nuclei of atoms were built of protons and neutrons, physicists have been trying to form a picture of the structure of the nucleus. The same task for the rest of the atom was completed in the first quarter of this century. We were able to understand in detail how the electrons move under the attraction of the nucleus, and how their motion is influenced by their mutual repulsion.

To achieve such an understanding requires three major steps: First, we must know the forces between the particles. Second, we need to know the mechanical laws which govern their motion under the influence of these forces. Third, we need in most cases a simplified picture, or model, from which to start. Once we have the first two ingredients, we could in principle write down a set of mathematical equations whose solutions would tell us all about the atom, or about the nucleus. In the simplest possible atoms, like that of hydrogen, in which there is only one electron, or in the simplest compound nuclei, like the deuteron, which contains only one proton and one neutron, such equations can be written down and solved without difficulty. However, for more complicated structures this head-on attack becomes much harder and soon exceeds the capacity even of modern electronic computers.

We are like men who encounter for the first time a complicated machine, and who try to analyze its operation. If we attempt, without any guidance, to puzzle out the interplay of all the parts of the machine, we should soon lose ourselves in a maze. Instead, we first try to ascertain the major features of the machine's operation. We then devise a model which resembles the real thing in these features, yet is simple enough to be analyzed. Then, of course, we must put in corrections for the complications which we have left out and check that they do not materially alter the picture.

In the study of the atom the first of the three steps hardly presented a problem. As soon as Ernest Rutherford had demonstrated that the atom consisted of a heavy, positively charged nucleus and of light, negatively charged electrons, it was taken for granted that the forces between them were the electric attraction of unlike charges, following the inverse-square law familiar to every student of physics. The major difficulty was the second step. It turned out that the basic mechanical principles of Isaac Newton, which applied to all "large" objects from the planets and the moon down to steam engines and watches, had to be revised in the atomic domain. To understand atoms we had to use the new ideas of the quantum theory, following the pioneer work of Niels Bohr, who adapted for this purpose the concept of the quantum of action which Max Planck had first found in the behavior of light. These new laws of mechanics were later formulated as the laws of "quantum mechanics," or "wave mechanics," which gave us complete command over the theory of the atom.

The third step, of finding a simplified model for discussing the atom, also proved relatively easy. In working out the possible orbits of a single electron under the attraction of a proton, as in the hydrogen atom, Bohr found that one could account for the behavior of a more complex atom by assuming that each of its electrons moved in such an orbit. The larger the number of electrons in an atom, however, the more distinct orbits they occupy; this is a consequence of the "exclusion principle" discovered by Wolfgang Pauli, which limits the number of electrons that can travel in a given orbit.

We must allow not only for the attraction of the electrons by the nucleus, but also for the repulsion of the electrons by one another. However, we simplify the nature of this repulsion by forgetting that it changes continuously as the electrons move around in their orbits, and treating it as a fixed field of force. In other words, we replace the repulsion due to a moving electron by that which we would obtain if the electron were spread out evenly over its orbit. This simplification can be justified by the fact that the repulsion acts over relatively long distances, so that each electron is at any time under the influence of several others. If we underestimate the effect of one of the electrons which may happen to be rather close to the one we are looking at, we are likely to overestimate the effect of another which happens to be rather far away.

This model of the atom is usually called the "shell model," because it is convenient to group together the electrons moving in orbits of similar size but of different shape and direction. Such a group of orbits is called a shell.

When the atomic nucleus first became an object of serious study, the nature of the difficulties was rather different. The general laws of dynamics did not seem to require further revision; the laws of quantum mechanics which had been discovered in atomic physics seemed quite adequate for the nuclear domain. Indeed, we have not yet found any evidence in the behavior of nuclei which would suggest that these laws might be in error. Thus the second step in our list presented no problem.

The Nuclear Forces

On the other hand the first step—the determination of the forces between the particles—proved to be a very difficult

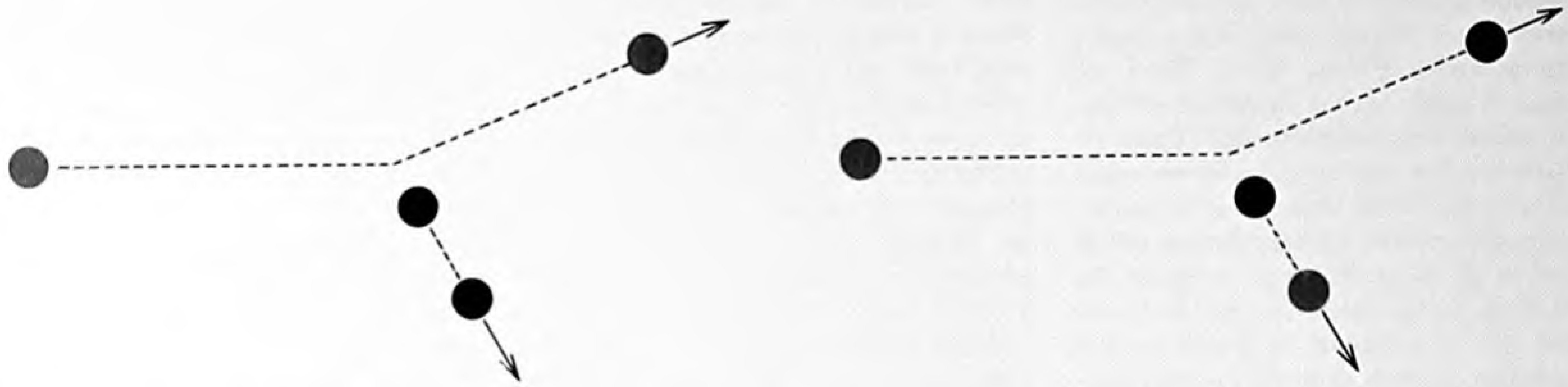
problem. Even today, after some 25 years of intense study, we cannot claim to have a complete answer, but we have by now at least a fair knowledge of what the forces are like.

They cannot be electric in origin. The only electric charges found in the nucleus are the positive charges of the protons, and like charges repel each other; thus electric forces cannot be responsible for holding a nucleus together. Moreover, electric forces are much too weak. We know that the energy of attraction of two unlike charges (*i.e.*, the work we have to do to pull them apart) varies inversely as their distance. The attractive energy of an electron and a proton in the hydrogen atom is a few electron volts (ev), and since the diameter of the hydrogen atom is 20,000 times larger than that of the smallest nucleus we should expect electric energies in the nucleus to amount to some tens of thousands of electron volts. Actually the forces inside a nucleus run to many million electron volts (mev). It follows that nuclear forces are vastly stronger than electric forces.

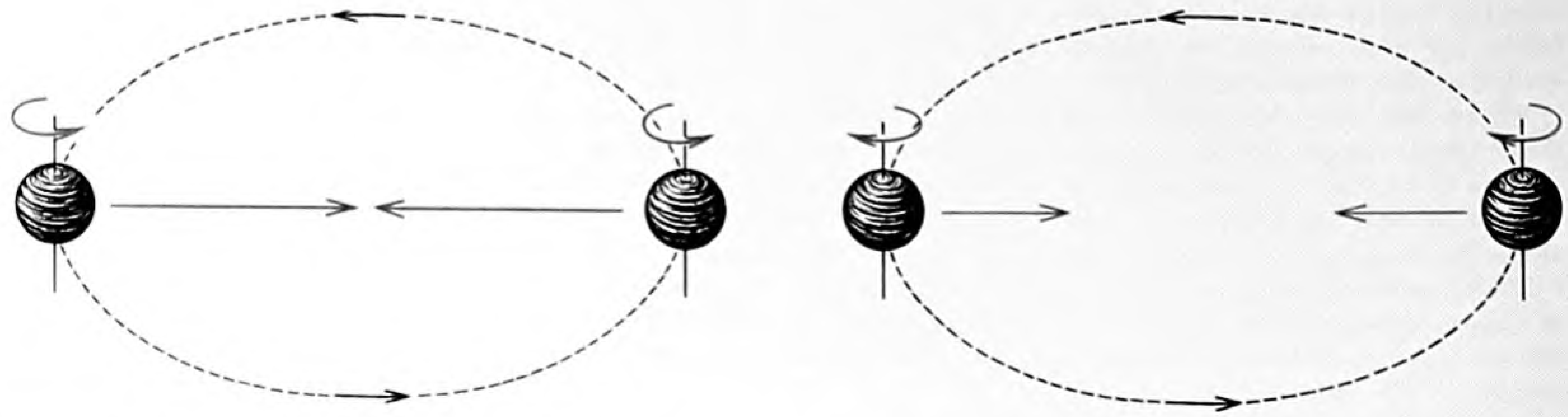
It is also clear that these strong forces act only over extremely short distances. The pioneer work of Rutherford on the passage of charged particles through matter showed that, even in encounters in which a charged particle approaches a nucleus to a distance of a few times the nuclear diameter, the only noticeable force is the electric one. We know today that nuclear forces between two particles are quite negligible if the distance between the particles is more than, say, four fermis. (The fermi, named for the late Enrico Fermi, is a convenient unit of distance for the nucleus. The diameter of a heavy nucleus is some 15 fermis; the diameter of the hydrogen atom, about 100,000 fermis.) It is not surprising, therefore, that earlier physicists did not meet nuclear forces in laboratory experiments. The only possible way of studying these forces is to observe the behavior of nuclei, or to bombard hydrogen or other nuclei with fast protons or neutrons under circumstances in which the effect of really close encounters can show up.

What makes this task harder is that the nature of nuclear forces, unlike the simple inverse-square law of electric or gravitational forces, is rather complicated. If the law of nuclear forces were simple, a few observations might suffice to guess its general form. But all simple guesses based on a few experiments have been disproved by later experiments. We are obliged to reconstruct the law of nuclear forces laboriously from the various pieces of evidence we can extract from the experiments.

Ultimately we hope to be able to derive the law of the forces from more basic principles, just as we can derive the inverse-square law of electric forces from the basic laws of electromagnetism. A beginning was made by the Japanese physicist Hideki Yukawa, who used the analogy with electromagnetic radiation to point out that nuclear forces must be related to a new form of radiation which could carry charged particles weighing a few hundred times more than the electron. His prediction was confirmed by the discovery of the so-called pi meson. His picture of the mechanism underlying the nuclear forces has been qualitatively confirmed by many observations, and has been a useful guide in our thinking about the forces. But it has not yet been possible to use his idea for a reliable and accurate derivation of the law of the forces because of the mathematical problems which stand in the way. We do not know today whether a correct solution of the equations em-



CHARGE EXCHANGE in the nucleus is schematically depicted. When protons *(black balls)* are struck by fast neutrons *(blue balls)* in half the cases (*left*) the neutron continues forward. In the other half (*right*), the proton exchanges its charge with the neutron.



SPIN-ORBIT FORCE arises from a relationship between spin and orbit. When two particles (*left*) spin in the same direction as that in which they move on an orbit, the force between them is strong. When they spin in opposite directions (*right*), force is weak.

bodying Yukawa's idea would yield the right forces, or whether there is something basically wrong with this approach. The difficulties arise chiefly from the greater strength of the nuclear forces, as compared to electric forces, which makes their mathematical analysis much more difficult.

Thus the best source of information about the forces still lies in direct experiments. These require collisions at high energies—much higher than the energies of particles inside ordinary nuclei. The reason for this is the wave aspect of particles, which is an essential feature of quantum mechanics. Slow particles are associated with waves of long wavelength, and collisions involving such slow particles do not provide much information about the finer features of the forces at work between them, just as in looking through a microscope at a dust particle with a diameter less than a wavelength of light we see only a general blur which does not reveal the shape or nature of the particle. To have particles of sufficiently short wavelength one must raise their energy to a few hundred mev. The most reliable information on nuclear forces has therefore become available only in the last few years, as a consequence of the development of accelerating machines which produce clean beams of protons, neutrons, or electrons with such energies. This need for high-energy beams is entirely similar to the situation in atomic physics, where detailed pictures of the structure of atoms require the use of X-ray or electron beams of several thousand ev—much greater than the energies of the electrons inside the atoms, whose wavelength is comparable to the atomic diameter. The complexity of the results has also made it necessary to call on the services of fast electronic computers for disentangling the observations.

I shall not attempt in this article to give anything like a complete specification of the nuclear forces, but shall stress only those features which are of importance for what follows. We have already noted that the forces must be strong and of short range. Since they hold the different particles together, they must on balance be attractive. At the same time they cannot be entirely attractive, since otherwise heavy nuclei would "collapse." By collapse we mean a state of affairs in which all the particles in a nucleus are so close together that each one is within the range of the attractive force of every other. In that case the attractive energy acting on each particle would grow with the total number of particles present, and the volume occupied by the whole nucleus would be the same no matter how many particles were in it. This is not found in reality. The energy per particle is roughly the same for all nuclei, light or heavy, and the volume of nuclei increases with the number of particles in them.

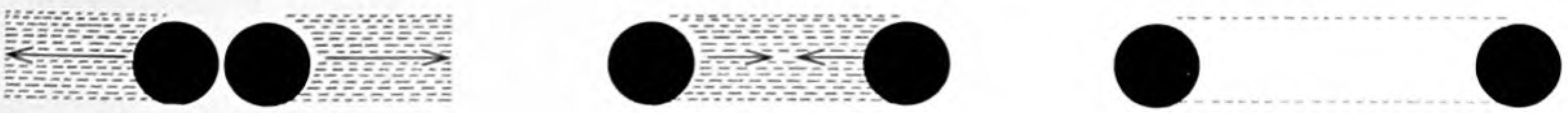
The Exchange Forces

This behavior, which indicates a limited attraction, is usually called "saturation" of the nuclear forces. There are two particularly plausible ideas to account for this saturation. One was suggested by the German physicist Werner Heisenberg, who was one of the founders of quantum mechanics. He postulated that at least part of the nuclear forces between a neutron and a proton involves an exchange of their position, so that after an encounter between them the neutron would tend to follow what had been the path of the proton, and vice versa. The exchange occurs readily only if the two move in very similar orbits, and, since the Pauli exclusion principle allows only a limited number of particles to follow the same orbit, such exchange forces would expose each particle to a strong attraction only from a few others. The bombardment of protons with fast neutrons confirmed this idea, because it showed that in most cases either the neutron or the proton tended to go forward with almost the same speed and direction with which the neutron had arrived. Since it is hard to deflect such fast particles from their path, this indicates that the incident neutron had continued almost in a straight line, but that in half the collisions it had changed its nature and become a proton, leaving a neutron behind.

However, the experiment also showed that only one half of the force was of the exchange type; the other half (corresponding to the neutrons still moving forward after collision) was an "ordinary" force. This is not enough to yield the required saturation, and some other factor must be involved. The second factor tending toward saturation is almost certainly a reversal of the direction of the nuclear forces at short distances, so that, as two particles approach each other, the attraction changes to repulsion. This concept of "repulsive cores" in the forces is familiar in the behavior of atoms. When atoms form chemical compounds, or liquid or solid substances, they are held together by attractive forces; but each atom has a fairly definite size, and when two atoms come into actual contact, their attraction changes into repulsion. We may liken this behavior to that of two rubber balls tied together with a rubber band. There is an attraction between the balls, but there is also a contact force which prevents the centers of the balls from approaching each other closer than one diameter. Shortly after the theoretical need for such a repulsive core in the nuclear forces had become clear, experiments on collisions between fast particles indeed showed direct evidence for these repulsive forces.

Among other features of the nucleus I should mention the "spin-orbit" force, that is, the dependence of the mutual interaction of two particles upon the direction of their orbit with respect to their spin. When the two particles spin on their axes in the same direction as that in which they revolve about each other, the attraction between them is stronger; when they spin in the opposite direction from that in which they revolve, the attraction is weaker. There is some evidence for such a spin-orbit force in experiments on nuclear collisions, but there is still some room for controversy in the interpretation of these experiments.

Our present knowledge of the nuclear forces, while still incomplete, is sufficient to discuss the behavior of nuclei and the collisions between them. At this point we meet the need for the third step in our general program, namely a simple model in terms of which we may approach the dynamical problem of the



NUCLEAR FORCES are dependent on the distance between particles. If the particles are very close, they repel each other (*left*). If they are a certain distance apart, they attract each other (*center*). If they are farther apart, they have little effect on each other (*right*).

motion of the 16 particles in the oxygen nucleus, or the 208 particles in the most stable lead nucleus.

Models of the Nucleus

The selection of a suitable model is not at all straightforward. Not that there is a shortage of suggestions. In fact the trouble in the recent past has been a surfeit of different models, each of them successful in explaining the behavior of nuclei in some situations, and each in apparent contradiction with other successful models or with our ideas about nuclear forces. In the past few years great progress has been made in bringing some order into this confusion and in understanding the justification for each of the models in the domain to which it is properly applied. I shall attempt to explain briefly some of the ideas behind these developments.

The most obvious idea was to use the shell model, which had been so successful in dealing with the atom. In fact, the first attempts to set up such a shell model were made even before the discovery of the neutron, when it was believed that nuclei were made of protons and electrons. A shell model with the wrong constituents cannot have much success in accounting for the facts, but in those days rather few facts were known, so such models were able to survive for some time.

After the discovery of the neutron, attempts to formulate a nuclear shell-model were renewed. This involved the idea of orbits (or quantum states) for the protons and neutrons, in which each of them was pictured as moving independentiy under the influence of some force which represented the average effect of the others, as in the case of the electrons in the atom. It did not seem possible, however, to choose groups of orbits of the right kind, so that the number of similar orbits which formed a shell could accommodate just the right number of neutrons and protons to account for the exceptional stability of nuclei with certain numbers ("magic numbers") of neutrons or protons.

The same idea was applied to the collision of neutrons with nuclei. According to the shell model, the impinging neutron should travel through the nucleus on its own orbit, as through some field of force, and individual encounters with the particles constituting the target nucleus ought to be rare and unimportant. Hence the neutron should in most cases emerge with the same speed as that with which it entered, and only rarely should it get trapped. The details of the process should not depend critically on the speed of the neutron.

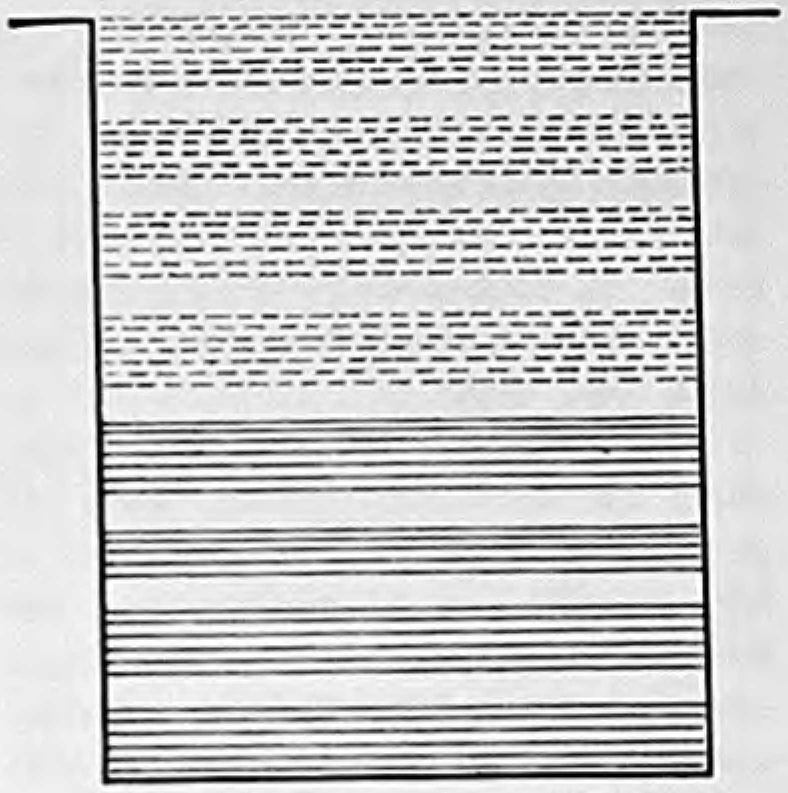
Observations of such collisions, initiated by Fermi in Rome, gave a completely different picture. Most of the neutrons that interacted with a nucleus were trapped, their excess energy being radiated in the form of gamma rays. Moreover, the chance of the neutron being affected by the nucleus depended very critically on its energy. One found a large number of resonances, *i.e.*, sharply selected energies, for which a neutron was sure to be picked up by the nucleus. For each target nucleus there are many such resonances, the energy difference between them being often as low as 100 ev, an exceedingly small difference on the nuclear scale.

These resonances turned out to be exceedingly sharp, and on the uncertainty principle of quantum mechanics a sharply defined energy is associated with a long time. So it follows that once a neutron gets into a nucleus in conditions of resonance it must stay there a long time —much longer than it would take it to cross a region the size of a nucleus.

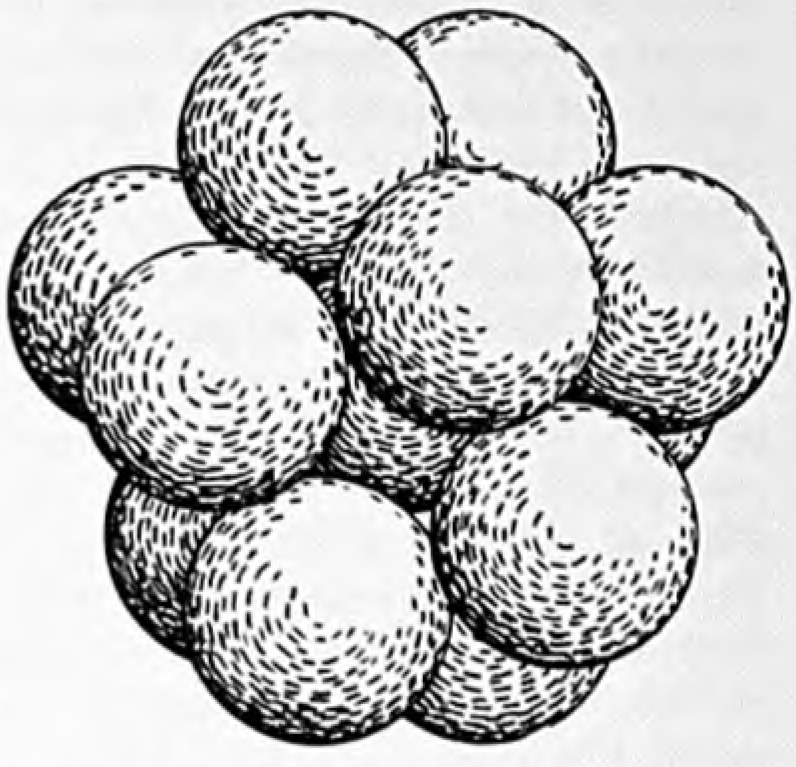
The Liquid-Drop Model

The way to resolve these apparent contradictions was pointed out by Bohr. He recognized that it was not right to think of a neutron as passing just through a general field of force, since the nucleus is densely packed with particles which each exert strong forces on the extra neutron as well as on each other. Instead of comparing the process with the passage of a comet through the solar system, as was appropriate for the passage of an electron through an atom, we should liken it to the entry of a golf ball into a space already fairly densely filled with similar balls. The result will be a complicated motion of all the balls, and the energy of motion of the extra one will rapidly get shared with the others.

The dynamical problem is now that of a true many-body motion, and we have vastly more possibilities of varying the details of the motion of all the particles. This means that the rules of quantum mechanics will give us far more states of motion, and these are responsible for the greatly increased number of resonances. We also see the reason for the long stay of the neutron in the nucleus, because when the energy of motion is shared among many particles, none of them can attain enough speed to escape from the general attraction. It must take a long time before by chance one of them col-

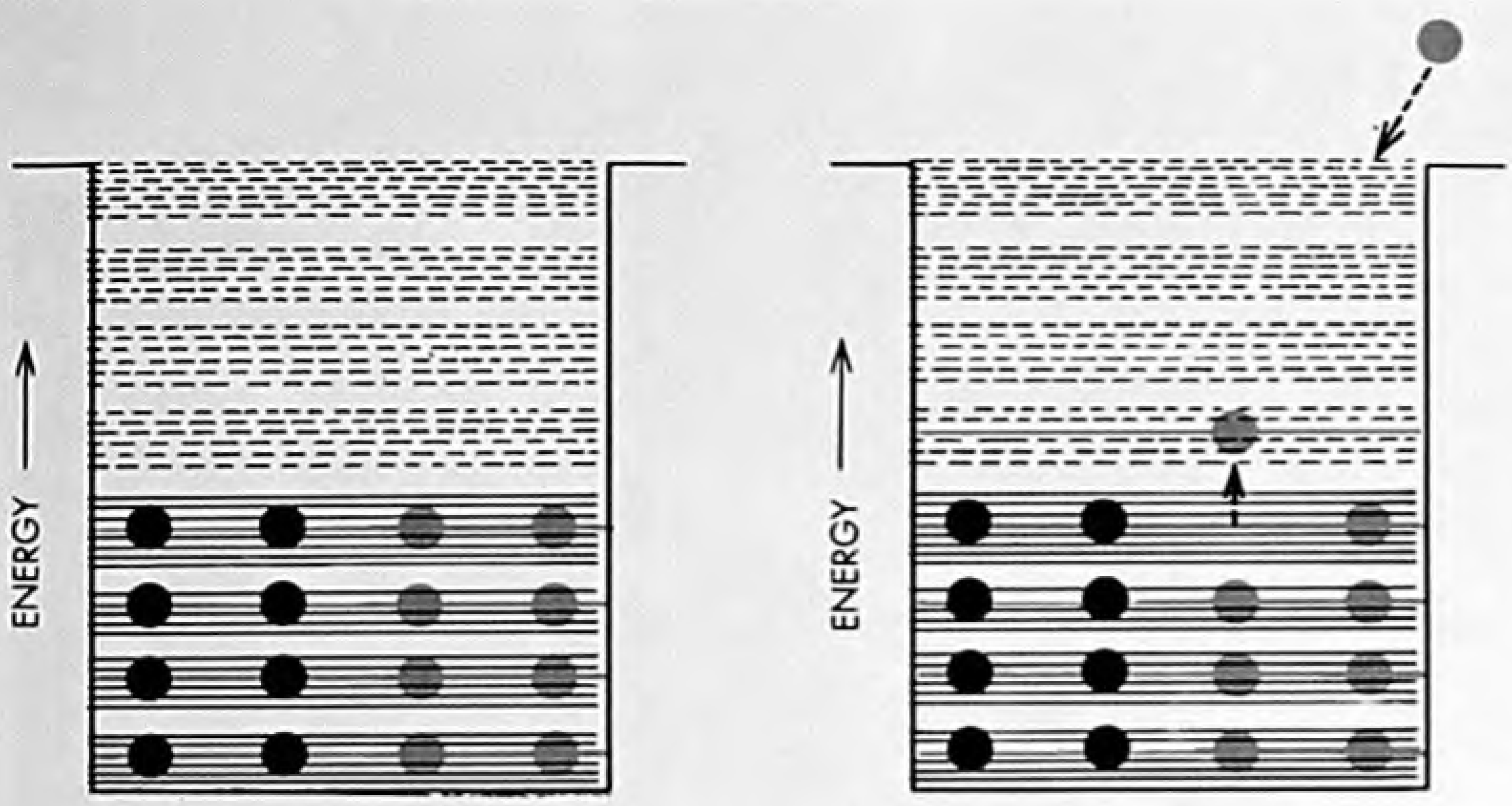


SHELL MODEL of the nucleus is represented by a potential "well" in which the groups of horizontal lines indicate orbits that can be occupied by particles in the nucleus. The groups of solid gray lines indicate orbits of lower energy; the groups of broken gray lines represent orbits of higher energy.



LIQUID-DROP MODEL may also be represented as a collection of golf balls. When another particle, or golf ball, enters the nucleus, the motion of all the balls is disturbed.

OPTICAL MODEL pictures the nucleus as a somewhat cloudy crystal ball. The cloudiness represents the tendency of bombarding neutrons to be absorbed by the nucleus.



LOW-ENERGY ORBITS in the shell model of the nucleus may each be occupied by only two neutrons (*colored balls*) and two protons (*black balls*). In the normal state of affairs (*left*) the low-energy orbits are filled; the particles cannot gain or lose energy, and thus cannot change their orbits. A bombarding particle (*upper right*) has energy to spare; thus it can exchange energy with a particle in nucleus and move it to orbit of higher energy.

lects enough of the available energy to get away. In our picture of the golf balls this will actually never happen, because in the meantime too much of the energy will have been dissipated in friction. In the nuclear case the analogue of friction is the loss of energy by gamma radiation, and this is responsible for the events in which the neutron gets trapped. But it is less effective than in the case of the golf balls, and some neutrons do get out again.

The physicist does not invoke here the similarity with a system of golf balls, which is not quite close enough, but he is reminded of a very similar situation which arises when a water molecule hits a drop of water, and for this reason Bohr's model is often called the "liquid-drop model."

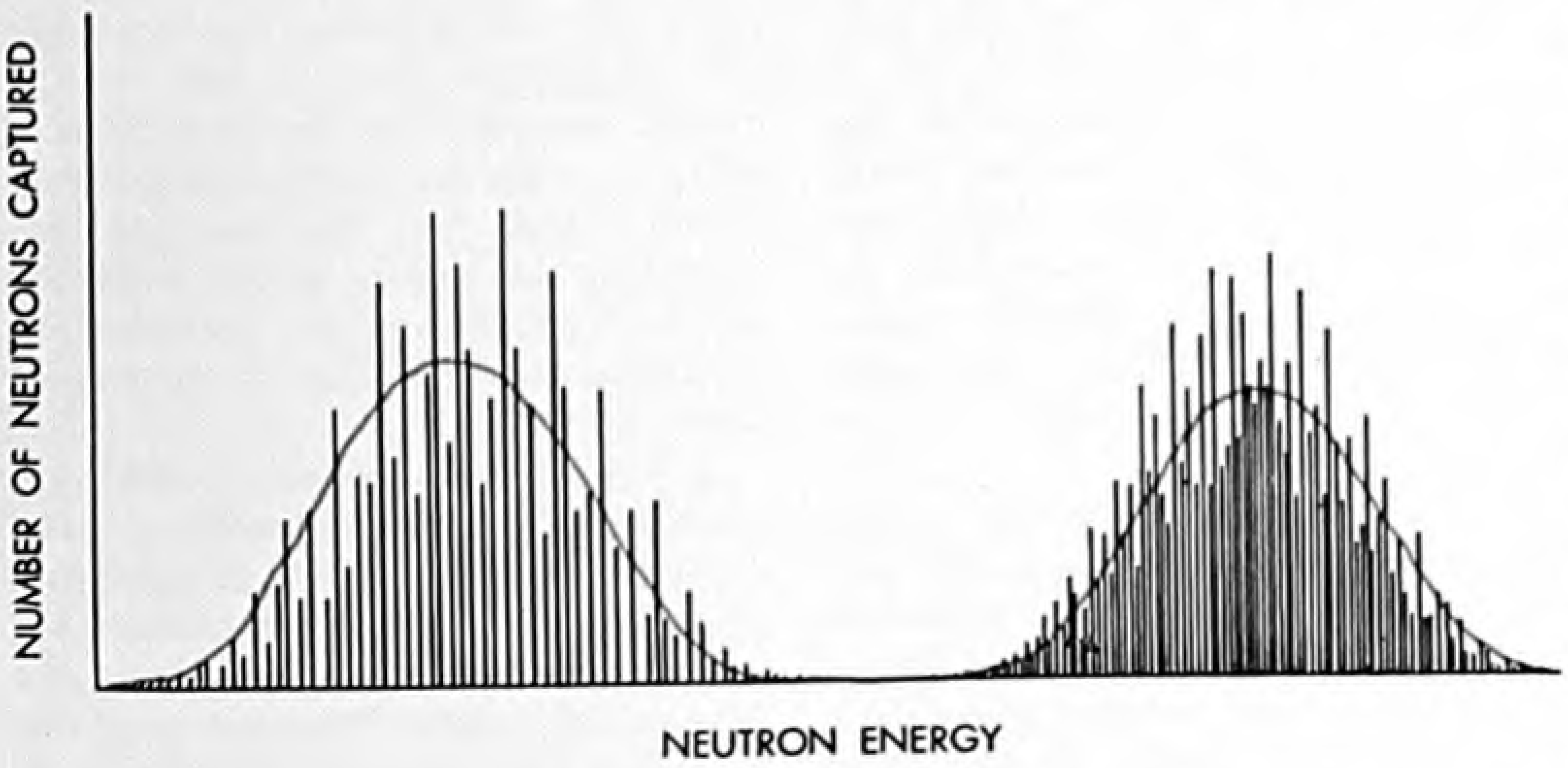
The liquid-drop model met with considerable success, and was able to explain many detailed features of nuclear reactions. At this time it seemed evident that the whole earlier idea of the shell model, which pictures the particles as moving independently, was doomed to failure, in view of the high density of the nucleus and the strong forces a particle was bound to experience in many encounters with others during the course of its motion. Most physicists then regarded the whole idea of a shell model as misconceived, but some, whether out of a stubborn refusal to accept the arguments against the model, or out of a deeper intuitive insight which convinced them that somehow one might be able to get around the argument, continued to look at the behavior of nuclei in their normal states in terms of shells.

The Shell Model Again

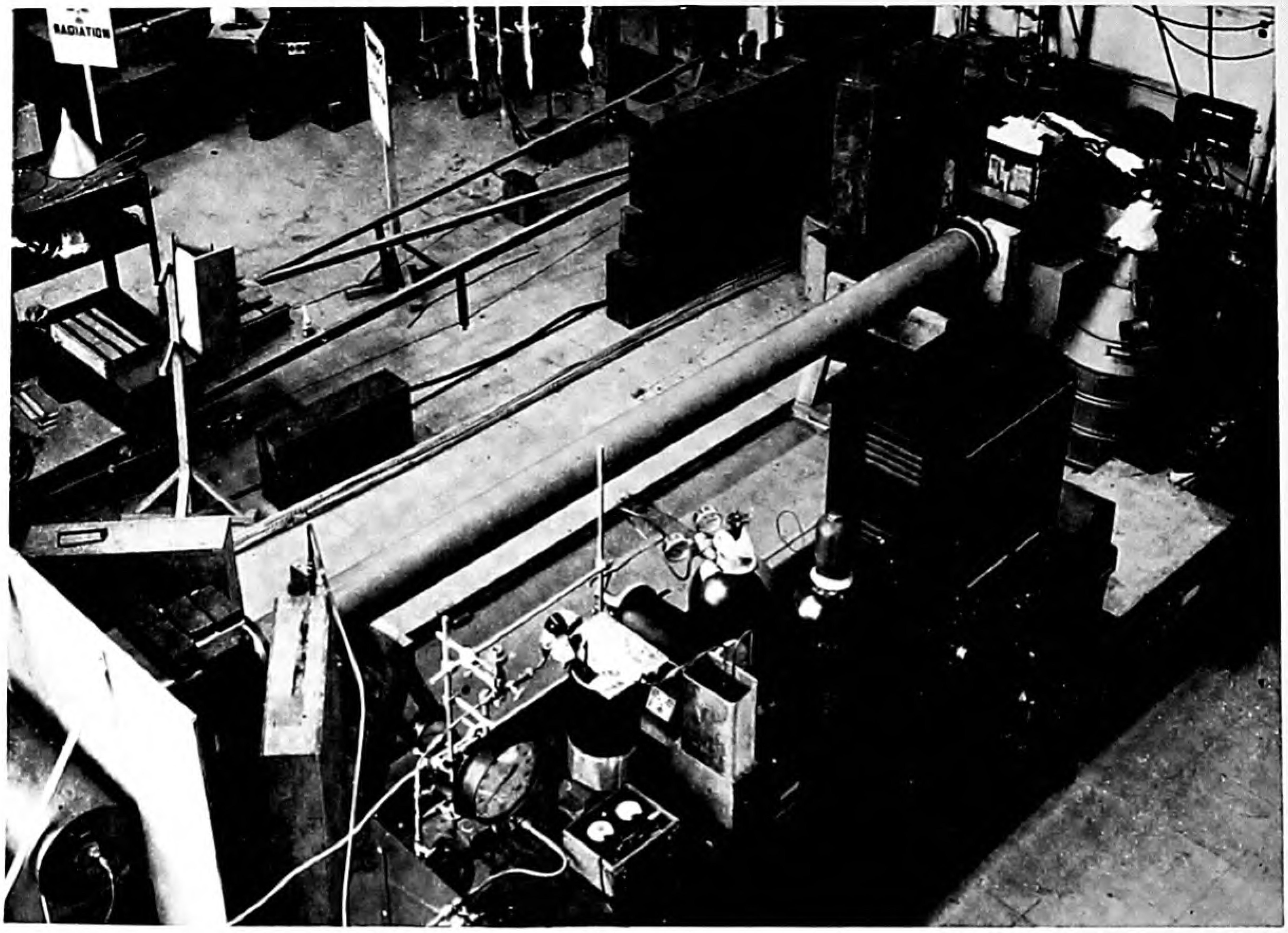
It soon became evident that there was overwhelming evidence in favor of such a shell picture, and the final success came when Maria G. Mayer of the University of Chicago and J. D. H. Jensen of Heidelberg independently noticed that the facts fitted amazingly well with a slightly modified shell model. The new feature was that when a particle spins in the direction in which it moves about the center of the nucleus, its orbit is different from the orbit of a particle spinning in the opposite direction. When this idea was put forward, it was not known that the force between two particles depends on the relative orientation of spin and orbit. Today the idea appears entirely natural. With this refinement, such a mass of data about the behavior of nuclei could be explained that there remained no doubt as to the essential of the particle being absorbed, *i.e.*, lost from the beam of bombarding neutrons [see "A Model of the Nucleus," by Victor F. Weisskopf and E. P. Rosenbaum; SCIENTIFIC AMERICAN Offprint 261]. How can we understand the success of this picture of independent particle motion in view of the Bohr argument?

The answer to this question has been given in essence by Weisskopf. It may be expressed by considering the time sequence of events. To be sure, the bombarding particle is likely to be disturbed from its path by collisions, but this will take a little time. So for a short time it will penetrate into the nucleus on a regular orbit, and this initial period is important for determining whether it will actually get deep inside or be turned back at the surface. Now, to recall once again the uncertainty principle, we know that in talking about a short time interval we must not try to specify the energy too accurately. We should therefore think not of neutrons with a well-defined energy, but of a beam of neutrons varying in energy by an amount that is greater the shorter the time in which they are likely to be involved in collisions inside the nucleus. Experiments often make use of such mixed beams, if the experimenter does not take trouble to select the neutron energies accurately. If we have data with accurate energy selection we should lump together the observations over a suitable range of energies.

Then we do not see the sharp resonances any more because there will al-



GIANT RESONANCES of a typical nucleus are indicated by the colored curve. Each of the vertical lines represents an ordinary resonance. The height of each line denotes the number of bombarding neutrons at that energy which are trapped within the nucleus, or which emerge from the nucleus with only part of their original energy. Giant resonances are observed when nucleus is bombarded with particles of lower energy and lower resolution.



OXYGEN NUCLEI ARE BOMBARDED with neutrons in this apparatus at the Brookhaven National Laboratory. The neutrons are produced by the Brookhaven nuclear reactor, the concrete shield of which is visible at right. The oxygen atoms are contained in the long tank in the middle of the picture. The neutrons which are not absorbed are counted in the shorter tank at lower left.

ways be many of them within the energy range we use. The result we get in this way will reflect the number and strength of the resonances within the selected range. But we may now think of these results also as determined by the first short time interval of the event, and as the neutron pursues a regular orbit during this short time interval the results now should reflect the behavior of such regular orbits. This therefore leads us directly to the picture of the optical model, which has neutrons traveling in regular orbits. The absorption which was allowed for in Weisskopf's optical model merely reflects the fact that the particles do not stay on such a regular orbit forever, but are sooner or later removed from it by collisions with other particles.

The strength of this absorption is thus related to the rate at which collisions occur inside the nucleus. If they are very frequent, so that the particle covers only a small fraction of the nuclear diameter before it hits something, the "giant resonances," which correspond to the orbits of a single particle, will become weaker and more diffuse. The fact that they are found to be pronounced and distinct shows that the particle has a fair chance of completing at least one revolution in its orbit. In this respect we see that the extreme form of Bohr's liquid-drop model, or our simple picture of golf balls, exaggerates the situation. But we have succeeded in reconciling Bohr's explanation of the many sharp resonances in terms of the many-body aspects of the problem, with the superimposed structure of giant resonances, which characterize the early stages of the process.

It remains to account for the quantitative features of the optical model—and in particular for the long time a particle can stay in its orbit before being thrown out of it by a close encounter with another particle—in terms of the basic forces. A promising attack on this problem is now under way. The workers engaged in it include G. E. Brown in the author's group at the University of Birmingham. In particular, the low rate of collisions is seen to be linked again with the effect of the exclusion principle. We have seen that this cuts down the rate of collisions in a normal nucleus drastically. In the impact problems where there is more energy to spare, the collisions are more frequent, because there are more orbits available that are not already occupied, but the prohibition is still partly effective and the collision rate is still a good deal less than that suggested by the picture of golf balls, for which all quantum effects, including the exclusion principle, are of no importance.

A picture thus emerges in which the various, apparently contradictory, models of the nucleus are seen as consistent parts of a whole, each appropriate for answering certain questions about the behavior of nuclei. There are problems for which yet other models have to be used, including the important "collective model" developed by Aage Bohr and B. Mottelson of Copenhagen, but it would exceed the scope of this article to describe them and show how they fit into the story.

The Author

R. E. PEIERLS is a German-born British physicist who worked during the last two years of World War II at the Los Alamos Scientific Laboratory. Born in Berlin, he divided his training (according to the European custom) among several universities: Berlin, Munich, Leipzig (where he studied with Werner Heisenberg) and the Swiss Federal Institute of Technology in Zurich. In addition, he paid frequent visits to the Copenhagen laboratory of Niels Bohr; these, he says, "contributed more to my development than any other contacts." After several years as Wolfgang Pauli's assistant in Zurich, Peierls journeyed on a Rockefeller Fellowship to Rome, where he studied with Enrico Fermi, and also traveled to Cambridge to study under P.A.M. Dirac. Upon the advent of the National Socialist regime in Germany, Peierls decided to remain in England where, since 1937, he has been a professor at the University of Birmingham.

Bibliography

A MODEL OF THE NUCLEUS. V. F. Weisskopf and E. P. Rosenbaum in *Scientific American*, Vol. 193, No. 6, pages 84-91; December, 1955.

THE STRUCTURE OF THE NUCLEUS. Maria G. Mayer in *Scientific American*, Vol. 184, No. 3, pages 22-34; March, 1951.

SCIENTIFIC AMERICAN December 1957, Vol. 197, No. 6, pp. 73-84 OFFPRINT 236

FUSION POWER

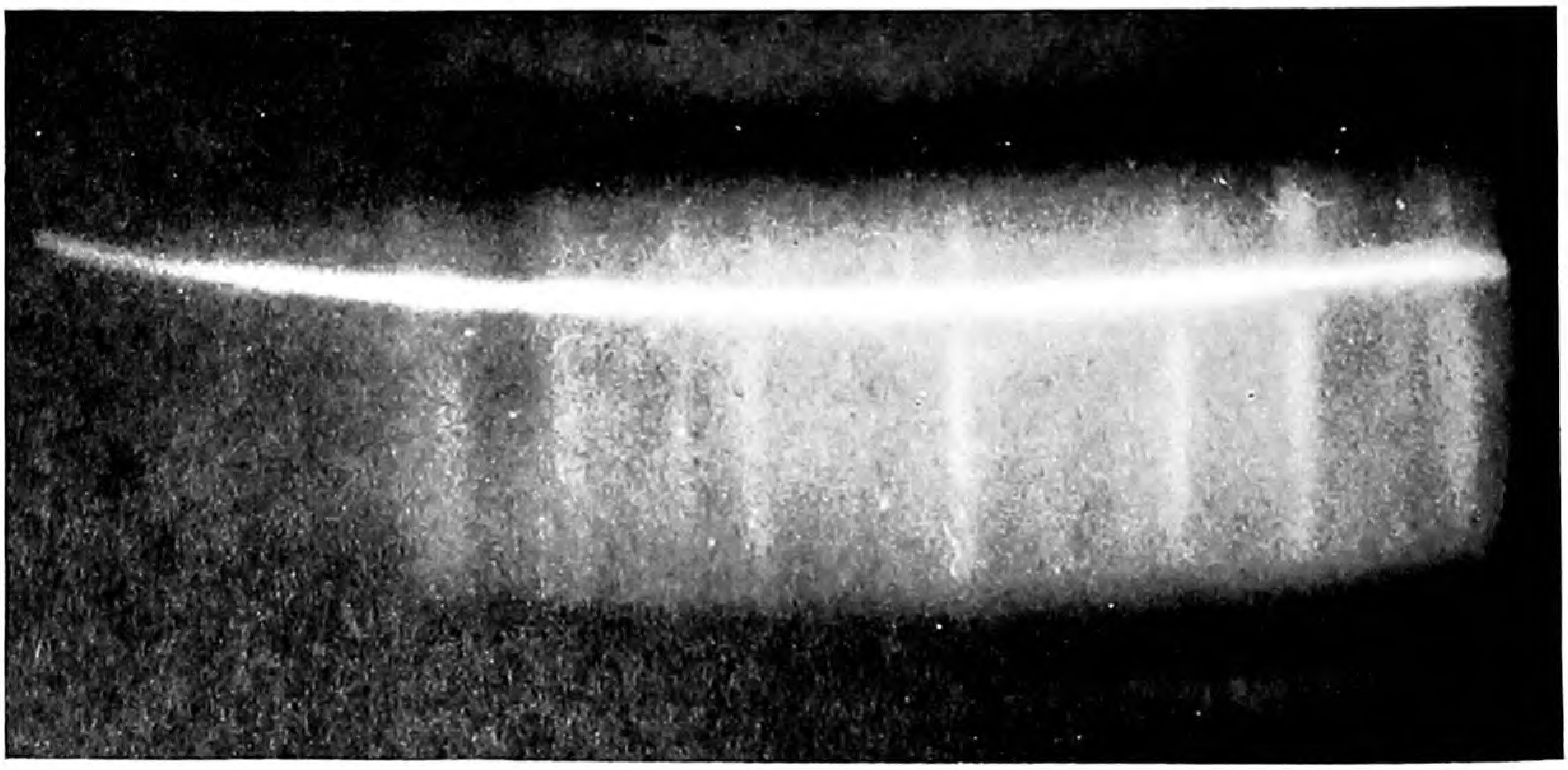
by Richard F. Post

If man can tame the reactions in which nuclei of atoms are fused rather than split, he will have an almost limitless source of energy. The problem is now being attacked in laboratories all over the world.

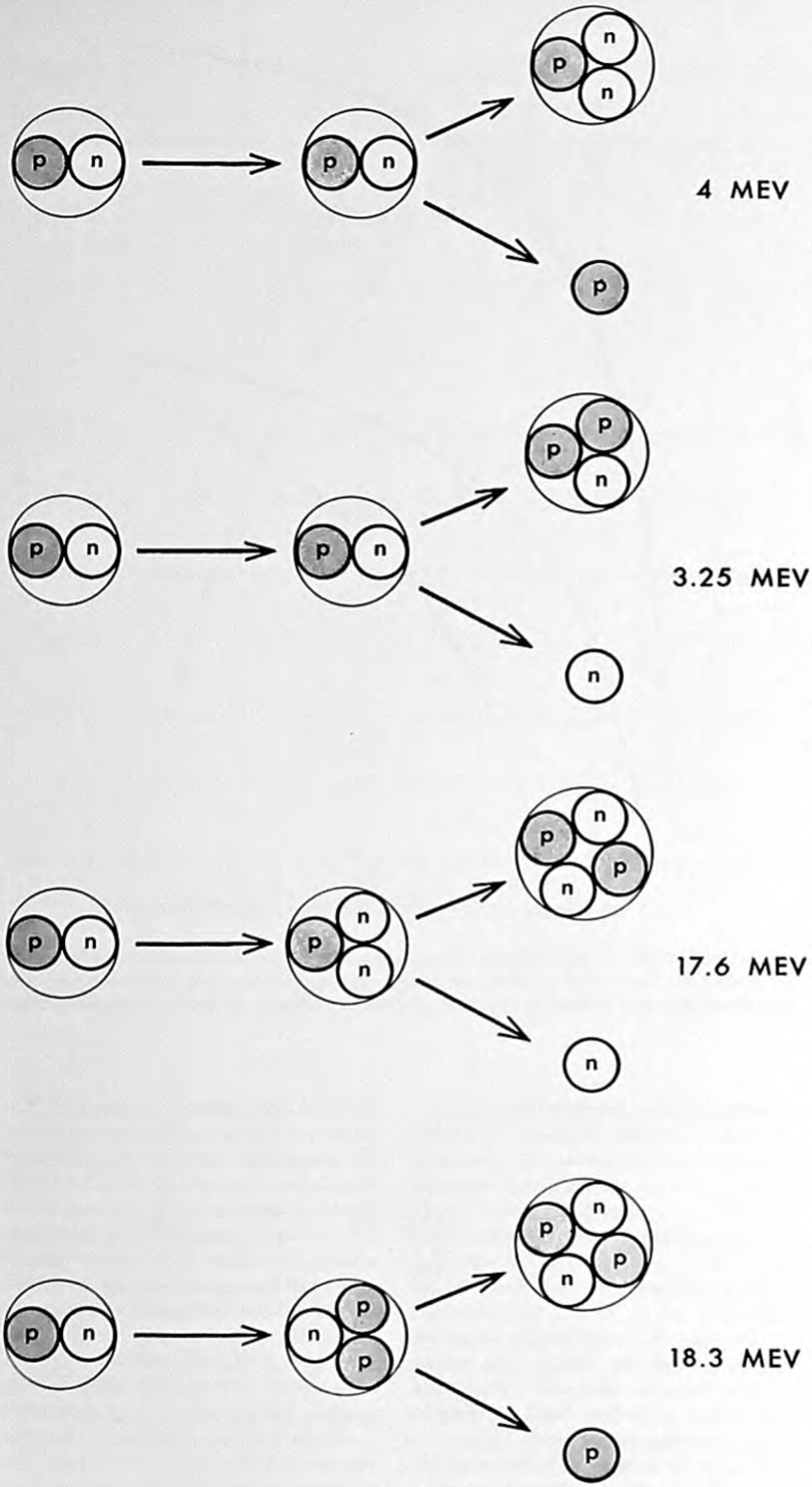
The effort to "tame" the hydrogen bomb reaction—to harness the energy of nuclear fusion for controlled power—is gaining momentum in many countries of the world. It is proceeding quietly, behind a screen of official secrecy, but nonetheless vigorously and on a substantial scale. Every nation has come to recognize that this research effort may well be the most important in the history of mankind. The ultimate stakes are so high, for nations individually and mankind collectively, that a growing sense of urgency and determination is infusing the work of the several nations on this problem.

The upsurge of our industrial civilization in the first half of the 20th century was founded upon fossil fuels—coal and oil. But there are many "have-not" nations, and even countries rich in these fuels are now seeing rapid inroads into their reserves. At this juncture uranium has come to the rescue as a hope for the future. Already electrical power is beginning to flow from the first nuclear fission plants. The United Kingdom expects to go over to fissionable fuels for most of its energy needs within a few decades, and many other countries are laying plans to follow suit; even the U. S., with its great reserves of coal, is spending hundreds of millions of dollars to prepare for turning to fission power at a not too distant date.

The world's uranium and thorium, it is estimated, represent an energy reserve somewhere between 10 and 100 times larger than its remaining coal. Even so, fissionable fuels, too, are an exhaustible supply. At the rate at which the world's energy needs are expanding (the U. S. has doubled its electrical power requirements every eight years) practically all of the economically recoverable uranium, as well as coal, might be exhausted within another century or so. Fission power also presents a more immediate problem: namely, disposal of its radioactive wastes. If the present power needs of the U. S. were all supplied by fission reactors, we would have to dispose each year of an amount of radioactive fission products equal to that from the explosion of 200,000 atomic bombs; by the year 2000, with increased use of power, the radioactive wastes



HOT PLASMA (an electrically neutral mixture of positive ions and electrons) is seen as a bright horizontal line in this image-amplifier photograph. The gas is heated and constricted by the "pinch effect," produced by a current of 8,000 amperes. This experiment, using krypton gas at a pressure of two ten-thousandths of an atmosphere, was done at the Los Alamos Scientific Laboratory.



FUSION REACTIONS which promise to be useful for production of power are shown schematically. At top two deuterons merge to form a tritium nucleus (triton). Second is the equally probable reaction in which the deuterons form helium 3 and a neutron. Third, a deuteron combines with a triton to form helium 4 and a neutron. At bottom a deuteron fuses with helium 3, producing helium 4 and a proton. The amount of energy released in each reaction is listed at its right. Protons and neutrons are designated as "p" and "n".

would come to the equivalent of eight million atomic bomb explosions per year! It is clear that the problem of safe disposal of radioactive ashes and gases in the coming age of fission power will soon become staggering.

All this helps to explain the drive, if not race, to find out whether thermonuclear power can be tapped and put to work. If the fusion reaction can be made to yield power, it will solve forever both the fuel supply problem and the problem of radioactive wastes. The basic fusion fuel, deuterium, is as inexhaustible as the oceans, and fusion produces no appreciable amount of radioactive byproducts.

The Fusion Reaction

Nuclear fusion is not exactly a new phenomenon. It has been generating the power of the sun and other stars for billions of years, and physicists discovered the fusion reaction in the laboratory before they did fission. But to create and control fusion power on the earth is a problem of a totally different order from harnessing fission. It is undoubtedly the most difficult project ever presented to scientists and engineers. This article is a report of the publishable progress made so far.

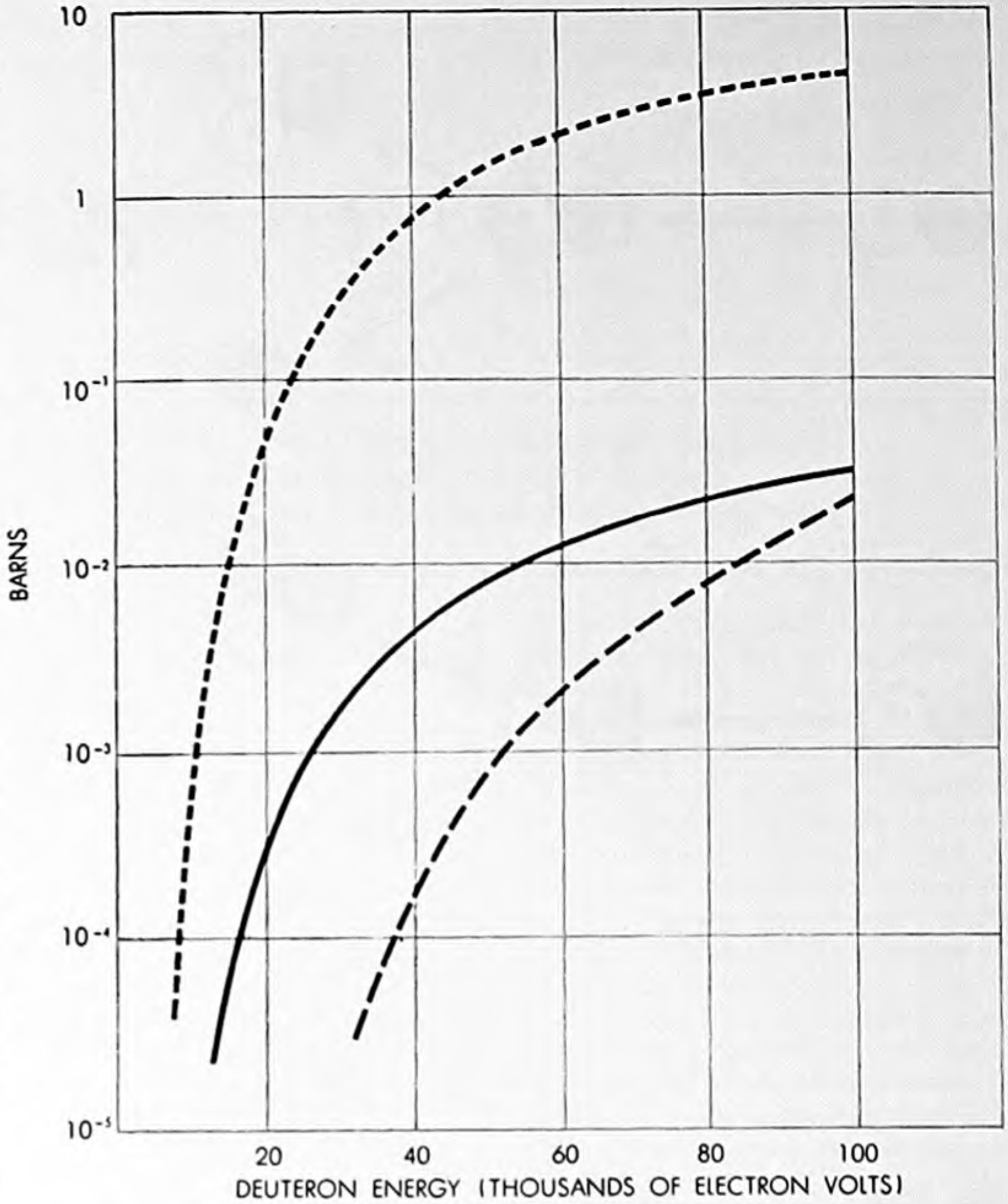
In the 1920s and 1930s physicists working with particle accelerators found that by accelerating protons (hydrogen nuclei) and other light nuclei to high enough energies (many thousands of electron volts) they could break through the nuclear electrical repulsion and force the projectiles to fuse with light nuclei in a target. The fusion releases energy, because part of the mass of the fusing nuclei is transformed into energy according to Einstein's famous equation $E=mc^2$. But in this sort of bombardment a great deal of energy has to be put in to make a few nuclei fuse. There can be no net yield of energy from fusion unless it proceeds by a self-sustaining reaction, as, for example, in the interior of the sun.

What is needed to produce a self-sustaining reaction? Here it is apropos to compare fusion with fission. The fission chain reaction is analogous to the explosion of TNT. A mechanical shock is sufficient to cause TNT to start exploding; the shock wave produced by fracture of its unstable molecules then touches off one molecule after another. Similarly in a fission chain-reaction the trigger for the successive fissions is supplied by neutrons, each fission releasing neutrons to attack more fissionable

uranium nuclei. In short, heat or kinetic energy plays no part in promoting the chain reaction.

A continuing fusion reaction, in contrast, is analogous to the familiar process of combustion. In ordinary burning, molecules combine (*e.g.*, hydrogen with oxygen, forming water), and their chemical reaction releases energy. To stick together or fuse, the molecules must collide violently, which means the material must be heated. Three conditions are needed to burn a chemical fuel and harness its heat to do work: (1) the fuel must be raised to its ignition point; (2) there must be enough of it to sustain a continuing reaction; (3) the energy released must be tapped in a controlled manner—*e.g.*, to heat water or to drive a piston. Now precisely the same conditions are needed to make a nuclear fusion reaction go and do useful work. The great difference is that for a fusion reaction the ignition point is rather high—hundreds of millions of degrees centigrade!

This one condition—the attainment of which was quite unthinkable on the earth until recently—underlies all our problems. From it stems a whole train of formidable questions. Unfortunately it seems that no one crucial experiment will tell us whether a solution is possible, as the first fission chain reaction did. We shall have to go through a long series of experiments to solve one problem after another. Most of them have to do with quantitative questions, that is, numbers. There never was a field in which the numbers were more imposing—or more important.



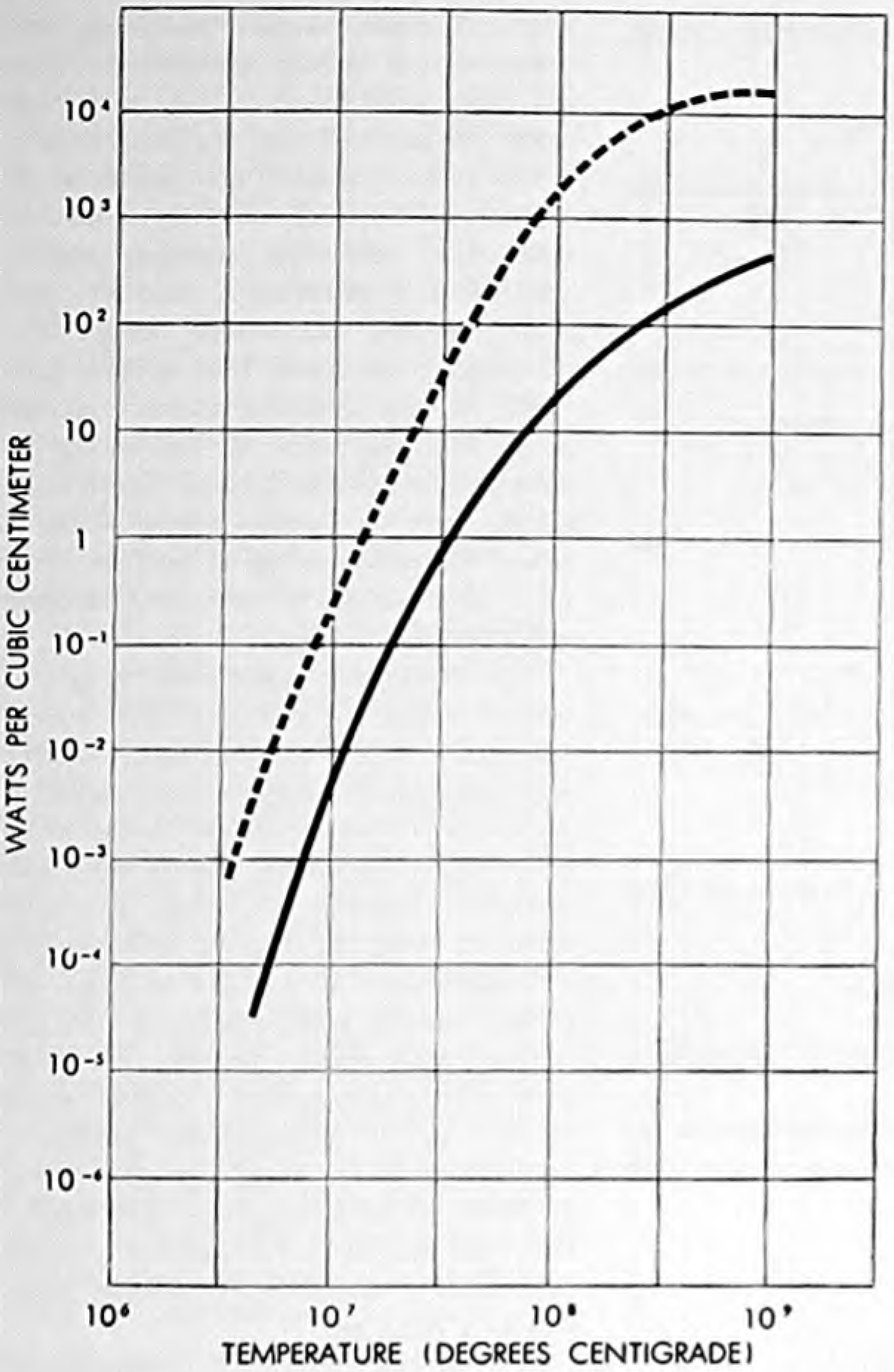
CROSS SECTIONS or probabilities of fusion reactions, as they vary with particle energy, are plotted for fusion of deuterium and tritium (*top*), deuterium and deuterium (*middle*) and deuterium and helium 3 (*bottom*). A barn is an area of 10^{-24} square centimeter.

The Fusion Fuel

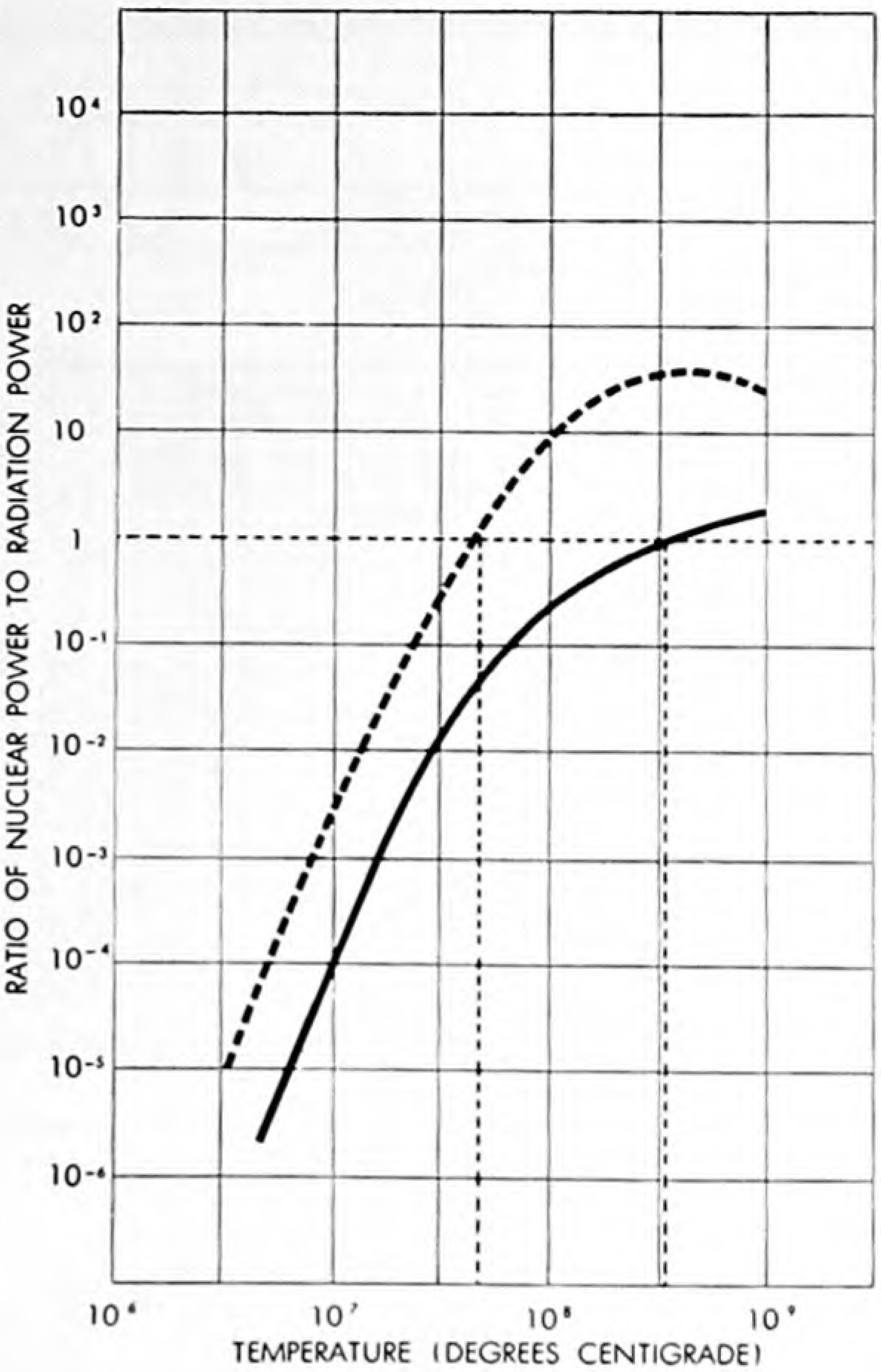
Our first concern is the fuel. The most interesting candidate is deuterium, a heavy isotope of hydrogen, found in ordinary water. The nucleus of the deuterium atom consists of one proton and one neutron. When two deuterium nuclei (deuterons) collide with enough energy to fuse, one deuteron grabs either the proton or the neutron of the other—the chances are 50-50. If it fuses with the proton (freeing the neutron), it forms helium 3, releasing about 3.25 million electron volts of energy. If it combines with the neutron (this time liberating its partner's proton), it yields about four million electron-volts and becomes hydrogen 3, or tritium, the radioactive isotope of hydrogen. A deuteron and a tritium nucleus (triton) will fuse more readily than two deuterons, and this reaction releases more energy. Tritium is therefore another potential fuel for thermonuclear reactors. It would have to be manufactured, however, as plutonium is, because there are only trace amounts of tritium in nature. One possible method of breeding tritium is to expose lithium to slow neutrons: on capturing a neutron, lithium splits into tritium and helium 4.

Complete "burning" of the deuterons and their products (tritium and helium 3) in a thermonuclear cycle would produce about seven mev (million electron volts) per deuteron burned. This corresponds to 43 million kilowatt-hours per pound of fuel. By comparison gasoline, one of the best chemical fuels, yields about six kilowatt-hours per pound. Although only a small fraction of the hydrogen in natural water is deuterium, still the deuterium in one gallon of ordinary water has an energy content equivalent to 350 gallons of gasoline! The oceans contain enough deuterium to supply the world with fuel for billions of years, even at a power demand 1,000 times the present figure. For use in our atomic energy program in the U. S., according to published figures, we are already producing an amount of deuterium which would be sufficient to supply the nation's total energy needs many times over if it could be burned as fuel. The cost of extracting deuterium from water is low enough so that deuterium would be less than 1 per cent as expensive as coal as a fuel. And finally, the nuclear burning of deuterium and tritium produces only inert gases, avoiding any problem of waste disposal.

Deuterium, then, represents the "ultimate fuel." But its great promise is matched by the equally great difficulty of finding a way to burn it!



FUSION-POWER OUTPUT of plasma at one ten-thousandth of atmospheric density is plotted against temperature. Upper curve is for deuterium-tritium fusion; lower curve, for deuterium-deuterium.



IGNITION TEMPERATURES, where nuclear power equals radiated power, are indicated by broken vertical lines for deuterium-tritium (*upper curve*) and deuterium-deuterium (*lower curve*).

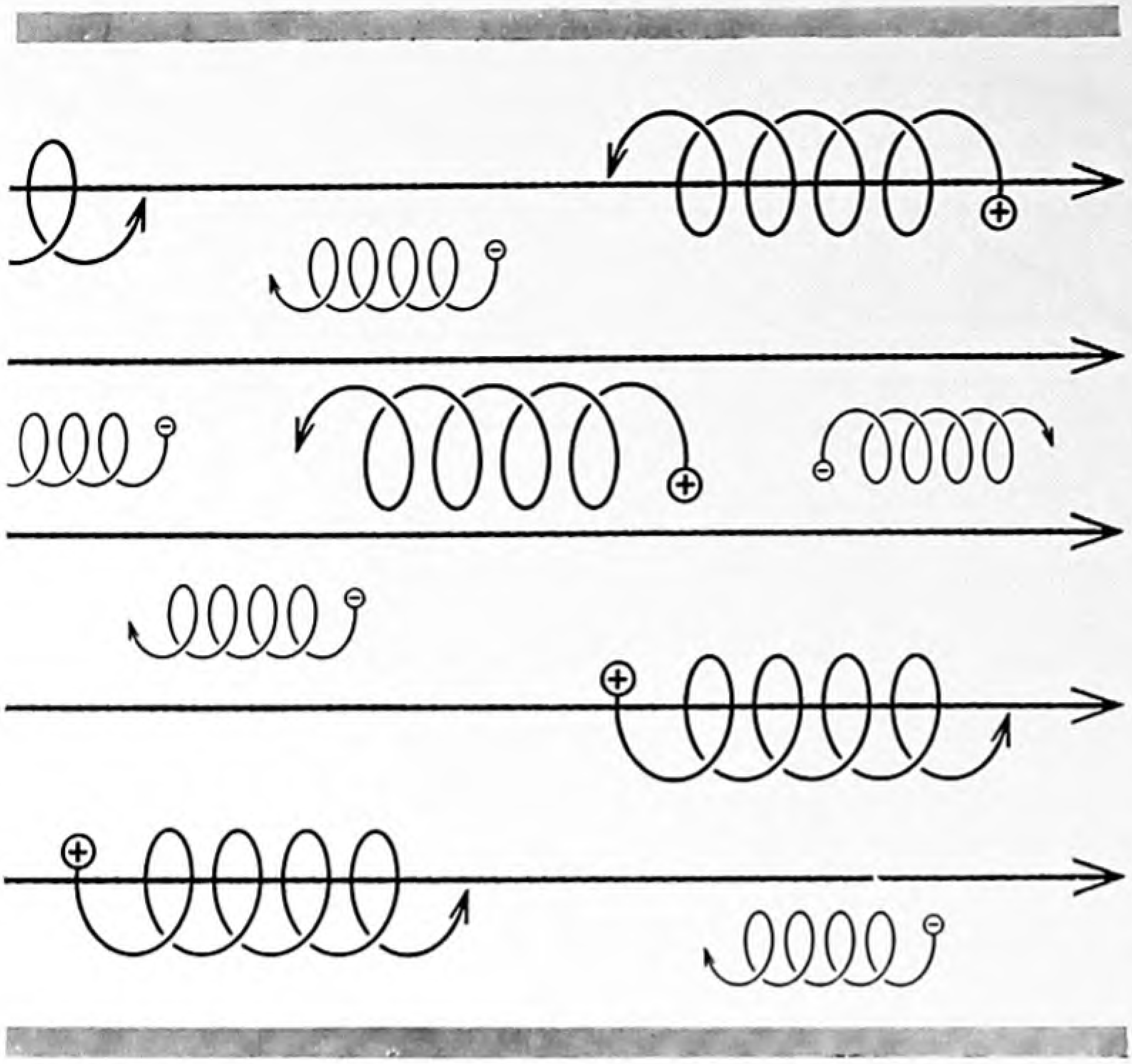
The burning of thermonuclear fuel in an uncontrolled manner has been achieved with all too well known success in the hydrogen bomb. But the problem of producing a controlled fusion reaction is quite different. Let us see what the necessary conditions are. The best way to do this is to follow an imaginary experiment.

The Plasma

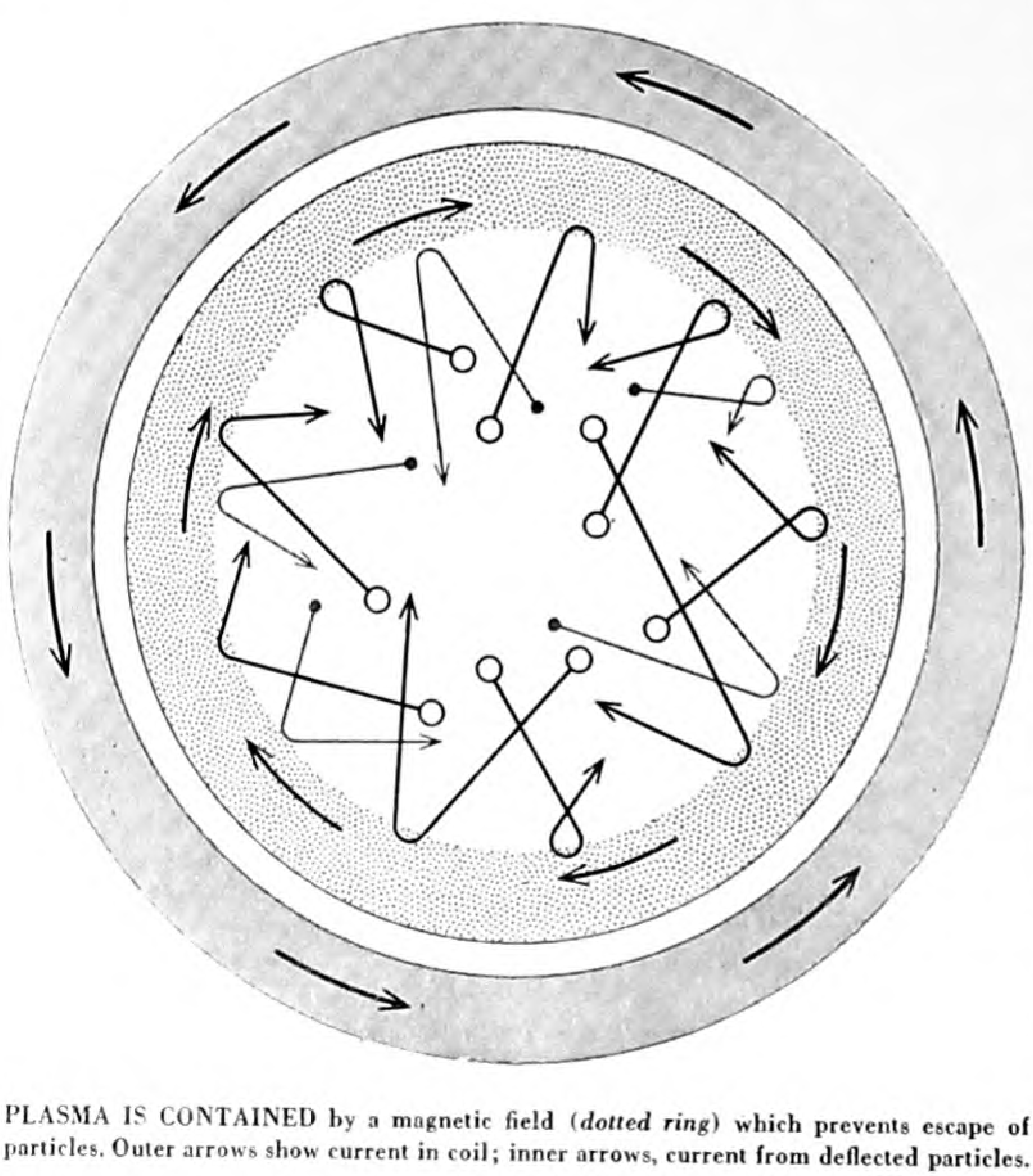
We take a liter of deuterium gas confined in a vessel made of a mythical material which is capable of withstanding the enormous temperatures and pressures that will arise in the course of the experiment. At room temperature and normal atmospheric pressure the deuterium gas-molecules are wandering about in the vessel with an average kinetic energy of about one 25th of an electron volt, or a velocity of about 3,000 miles per hour. Of course no fusion reactions are taking place. Now we heat the gas to 5,000 degrees C. At this temperature we no longer have molecules: the violence of their collisions has broken them apart into deuterium atoms. The pressure has risen to about 40 atmospheres (600 pounds per square inch), and the average velocity of the atoms is about 40,000 miles per hour. But we are still very far from the velocity needed to make two nuclei fuse.

Next, let us jump to 100,000 degrees. The remarkable properties of the mythical wall material are very much needed now, for any real material would long since have vaporized. Now the deuterium atoms of the gas have been broken down to the electrically charged nuclei (deuterons) and electrons: in a word, the gas has become what is known as a plasma. The gas pressure has risen to 1,500 atmospheres. The average velocity of the electrons is 10 million miles per hour, and even the much heavier deuterons are moving at the great speed of 170,000 miles per hour. Yet the deuterons still do not have sufficient energy effectively to overcome their mutual electrostatic repulsion. At this temperature there would be only about one fusion in the liter of plasma every 500 years! We still have a long way to go before we shall reach the ignition temperature of a mass of deuterons.

At one million degrees, the rate of fusion reactions will increase more than a billion billion times, but the total energy output will still be too small to be detected—only a few millionths of a watt per cubic centimeter. At 100 million degrees, however, the reaction rate will become really respectable. The pressure



PATHS OF PLASMA PARTICLES in a magnetic field are helixes. Positive particles go counterclockwise; negative particles, clockwise looking in the direction of the field.



PLASMA IS CONTAINED by a magnetic field (*dotted ring*) which prevents escape of particles. Outer arrows show current in coil; inner arrows, current from deflected particles.

then will have reached the staggering value of 1.5 million atmospheres. The electrons will be traveling at 90,000 miles per *second*, and the deuterons at 1,500 miles (around the world in 16 seconds). Essentially all of the deuterons will react with one another rapidly (within a fraction of a second), and their reactions will release energy at a fantastic rate—about 100 million kilowatts. But we shall not yet have arrived at the kindling point: to sustain the reaction we shall still have to put in more energy than the fusions release. Only at about 350 million degrees will the "fire" (*i.e.*, thermonuclear reaction) become self-sustaining.

This imaginary experiment brings out several important points. First, we need extremely high temperatures, though when we speak of high temperature here we are not thinking of heat in the usual sense but of the kinetic energy of the gas particles. Second, we could not even think of using the fuel at ordinary gas concentrations. If we are to keep the energy output and pressure of the gas within controllable bounds, we must start with a thin gas at a density much lower than at atmospheric pressure—somewhere in the neighborhood of one 10,000th of an atmosphere. But what a thin fuel this is! In a laboratory a gas at this density would be considered practically a vacuum.

One of the interesting consequences of using a very low density is that even though the plasma is very hot in terms of the speed of its particles, its heat *content* will actually be very small. A liter of deuterium plasma one 10,000th of an atmosphere in density would, at a kinetic temperature of 350 million degrees, have a heat content amounting to 18,000 calories—about enough to heat a small cup of coffee.

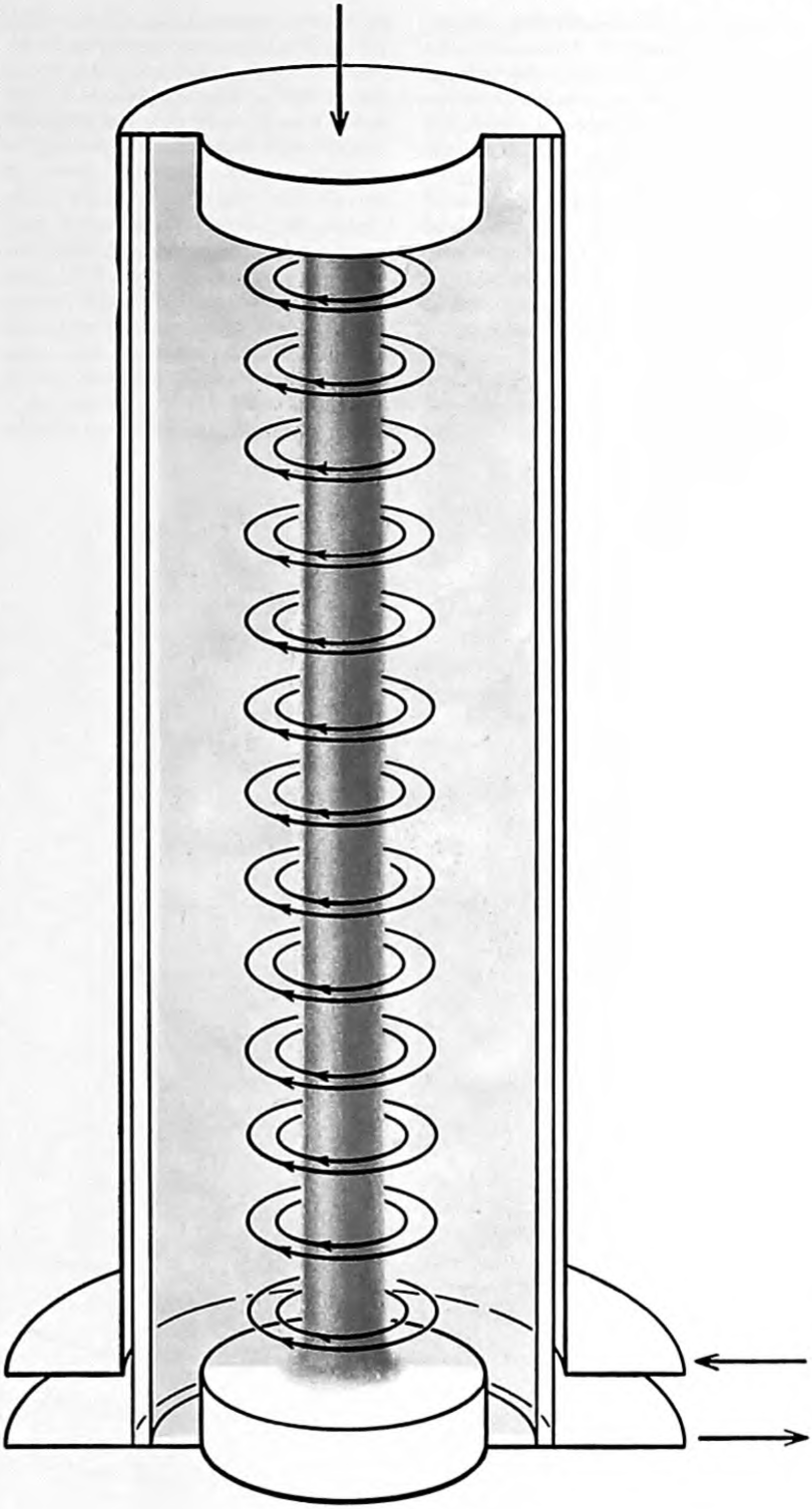
To calculate the rate of fusion power production from a hot plasma it is only necessary to know the reaction cross sections and to insert these as data in the theory of thermonuclear reactions in a hot gas [*see charts on page 297*]. However, calculations of this kind, important as they are in specifying required physical conditions, shed no light on *how* to heat a gas to thermonuclear temperatures or on whether the reactions, once initiated, could be made self-sustaining. This latter question depends on how much of the energy will be lost by radiation and other mechanisms. In the sun, the energy generated within its huge volume is sufficient to maintain the reactions in spite of the radiation loss from the surface. If we could build a

fusion reactor as big as the moon, we would not need to worry particularly about energy losses. But for a reactor of practicable size this is our key problem.

Let me first mention the unavoidable losses that must be lived with. These are the losses by radiation from the plasma—primarily in the form of X-rays emitted when electrons collide with nuclei. Now at a temperature of 100 million degrees, one cubic centimeter of dense matter would radiate energy at the unbelievable rate of three million million million kilowatts! But fortunately the rate of radiation drops very rapidly as the density of the matter falls. At the low density we have been considering, the radiation loss becomes comparatively small. The energy yield of fusion reactions increases rapidly with rise in temperature, and it will outstrip radiation losses at a temperature above 50 million degrees in the case of the deuteron-triton reaction and above 370 million degrees in the deuteron-deuteron reaction. These, then, are the ignition temperatures of the respective fuels. But the plasma must be very pure, because nuclei of higher elements (above hydrogen and helium) greatly accelerate the rate of radiation. A surprisingly small amount of impurities could poison a very large volume of plasma. For example, the metal in the head of a pin, if vaporized, would be quite sufficient to poison several railroad tank cars full of plasma. It is clear that purity will be a prime requirement in any controlled-fusion reactor.

Of the other class of energy losses—the ones we can and must reduce—the most serious is dissipation of particle energy to the walls of the reactor. We have to have a closed chamber, to hold out the atmosphere and keep our gas at low density. But consider the particles in this gas. They are so widely dispersed in our near-vacuum that each deuteron, in its random wanderings, travels thousands of miles on the average before it encounters another deuteron (or a triton). It has a far greater chance of hitting the walls of the container first. Yet if it does, this collision will immediately damp its energy. Obviously we cannot allow the particles of the plasma to touch the walls. Contrary to a common impression, the reason is not that the plasma will vaporize the walls (it does not contain much heat) but simply that contact with the walls would instantly cool the plasma and quench the reaction.

This, then, is the nub of the problem: How to confine a very hot gas within a material chamber (for at least a fraction



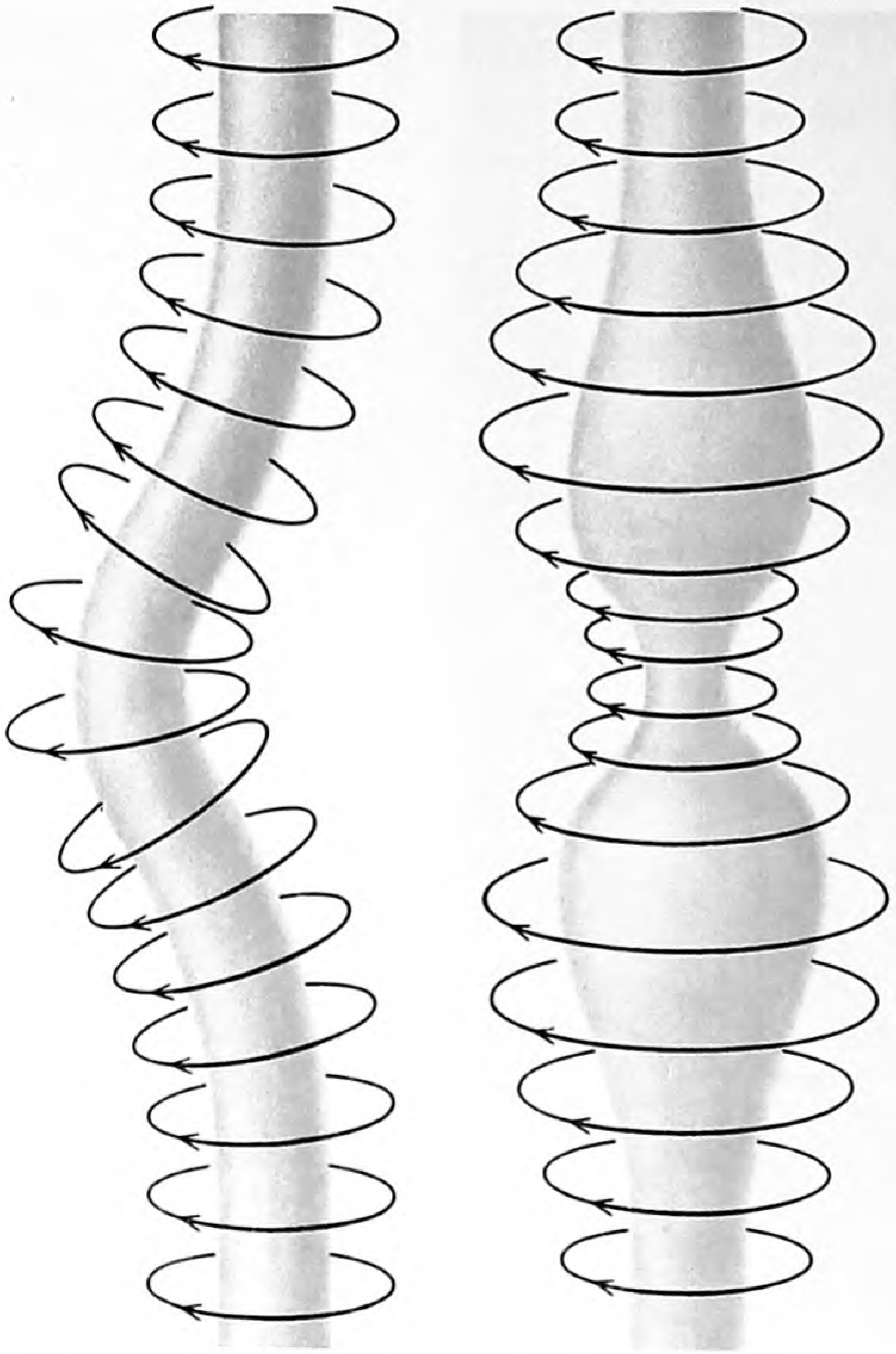
PINCH EFFECT occurs when a large electric current is sent through a plasma in a cylinder. Circular magnetic lines of force set up by the current contract and pinch the plasma into a narrow channel (*colored column*). The straight arrows show the direction of the current.

of a second) without allowing any appreciable amount of it to reach the chamber walls. Posed in this way, it sounds like a science-fiction problem, quite unsolvable in any real world. But as is now well known, about a decade ago an ingenious solution emerged—namely, the plasma might be confined within a magnetic field, serving as a kind of furnace liner in the chamber to keep the particles away from the walls.

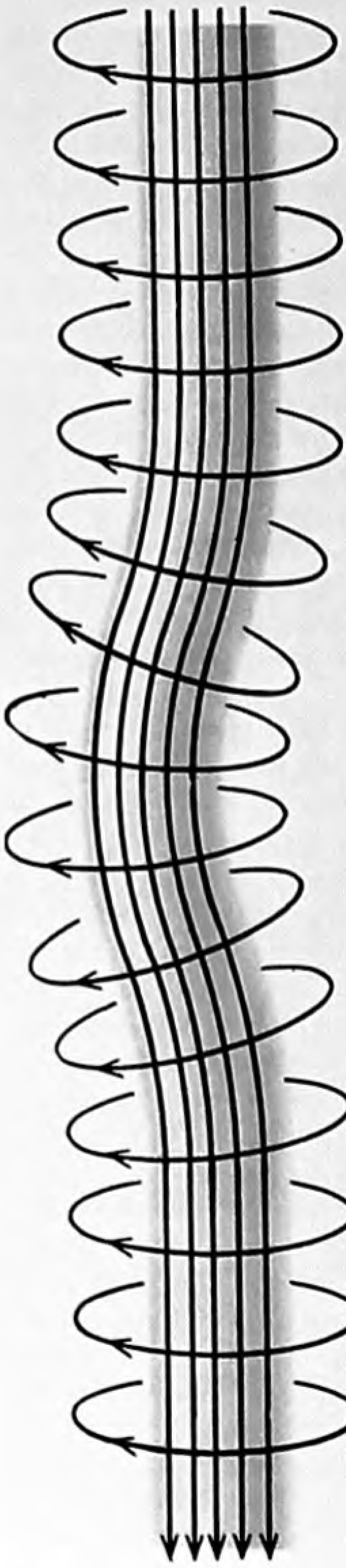
The Magnetic Bottle

The idea rests basically on the simple fact that a strong magnetic field will deflect charged particles from a straight path [*see diagram at top of page* 298]. Now a hot, high-pressure plasma could, under the proper circumstances, generate an internal magnetic field of its own strong enough to exclude the externally applied field. Inside such a plasma the particles would therefore move in straight lines. But at the boundary of the plasma they would be deflected back into it by the outside magnetic field [*see diagram at bottom of page* 298]. The magnetic lines of force, acting like elastic rubber bands, could resist considerable pressure. If the magnetic field were made strong enough, it should form a magnetic "wall" able to contain a high-pressure plasma, just as a steel cylinder holds a high-pressure gas. According to the theoretical calculations, a field with a strength of 50,000 gauss, for example, could withstand a plasma pressure of 100 atmospheres, and a field 10 times stronger (which has been achieved in laboratories) could support a pressure of 10,000 atmospheres.

A fusion reaction sustained in such a magnetic bottle could never "run away," as the fission chain reaction may. If the plasma pressure became stronger than the magnetic field, it would rupture the magnetic wall and the plasma would touch the material chamber wall, which would immediately quench the fusion reaction. By the very nature of the beast,



INSTABILITIES in a pinched plasma arise from kinks (*left*), where the magnetic lines are crowded on the concave side, and constrictions (*right*), where the field is also crowded.



STABILIZED PINCH might be achieved by putting a magnetic field (*long arrows*)

then, a fusion reactor could never explode; it could only collapse.

The simplified picture we have been considering should not be taken to mean that the magnetic bottle would be leakproof. Actually the plasma and the magnetic field would gradually penetrate and intermingle with each other, and the plasma would eventually escape completely unless replenished. Fortunately this leakage should be slow enough, according to the theory, to permit the achievement of a self-sustaining fusion reaction. But, as we shall see, a magnetic bottle does not always behave as the simple theory predicts. The interactions between a high-temperature plasma and magnetic fields are a difficult problem in fundamental physics, and they have given rise to a new field of study which might be called "experimental astrophysics."

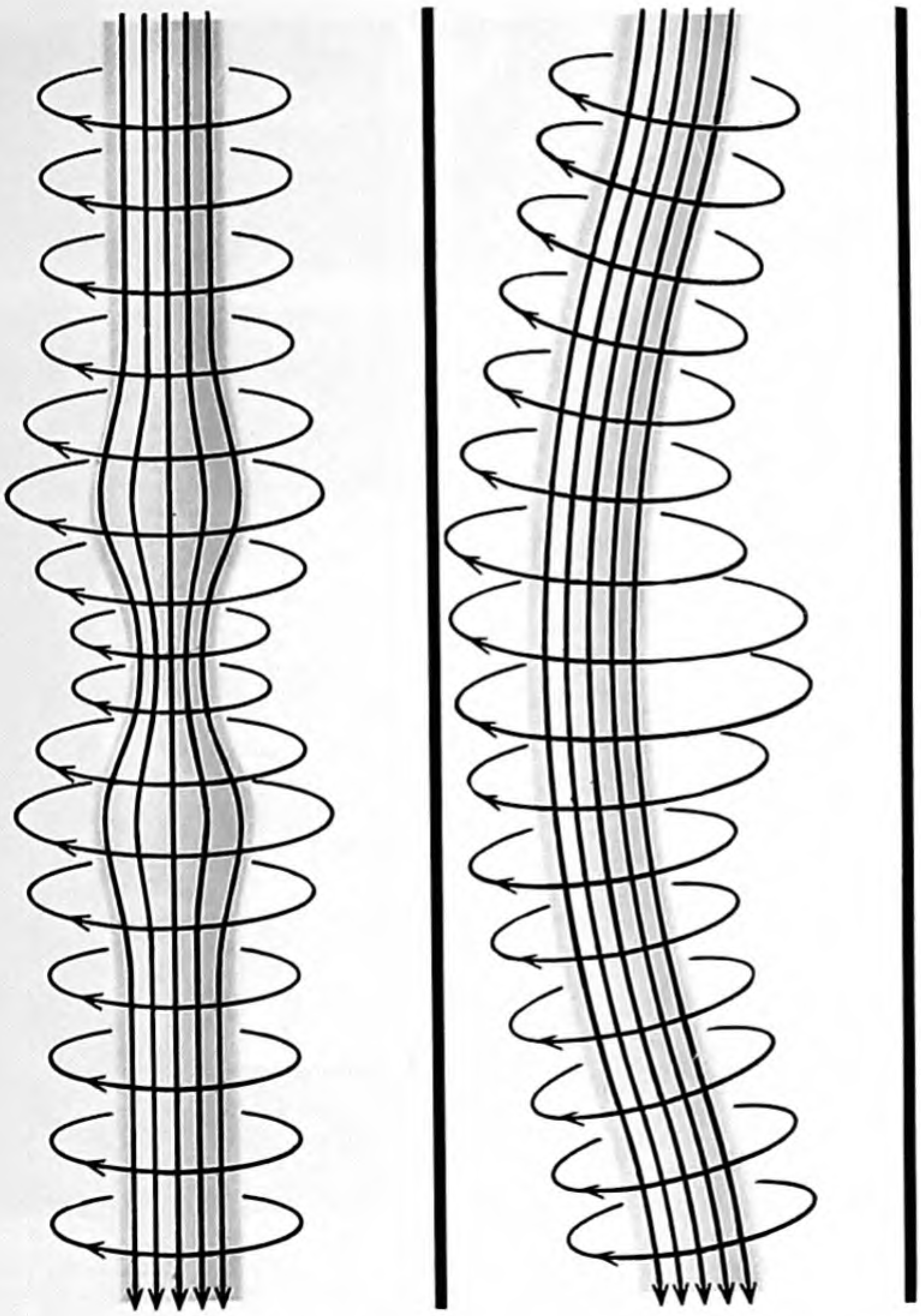
The Pinch Effect

So much for the theory. How could we actually make a magnetic bottle? It occurred independently to many investigators that an obscure electrical phenomenon known as the "pinch effect" might provide the answer. This effect, first produced experimentally only about a decade ago, requires a very large electrical current. When such a current is passed through a conducting gas in a tube, it sets up a magnetic field which tends to pinch the gas and pull it away from the tube walls [*see diagram on page 299*]. Magnetic lines of force circling the gas compress it by their tension. Since a plasma is an excellent conductor of electricity, the pinch effect looked like an attractive and ready-made means of forming a magnetic bottle.

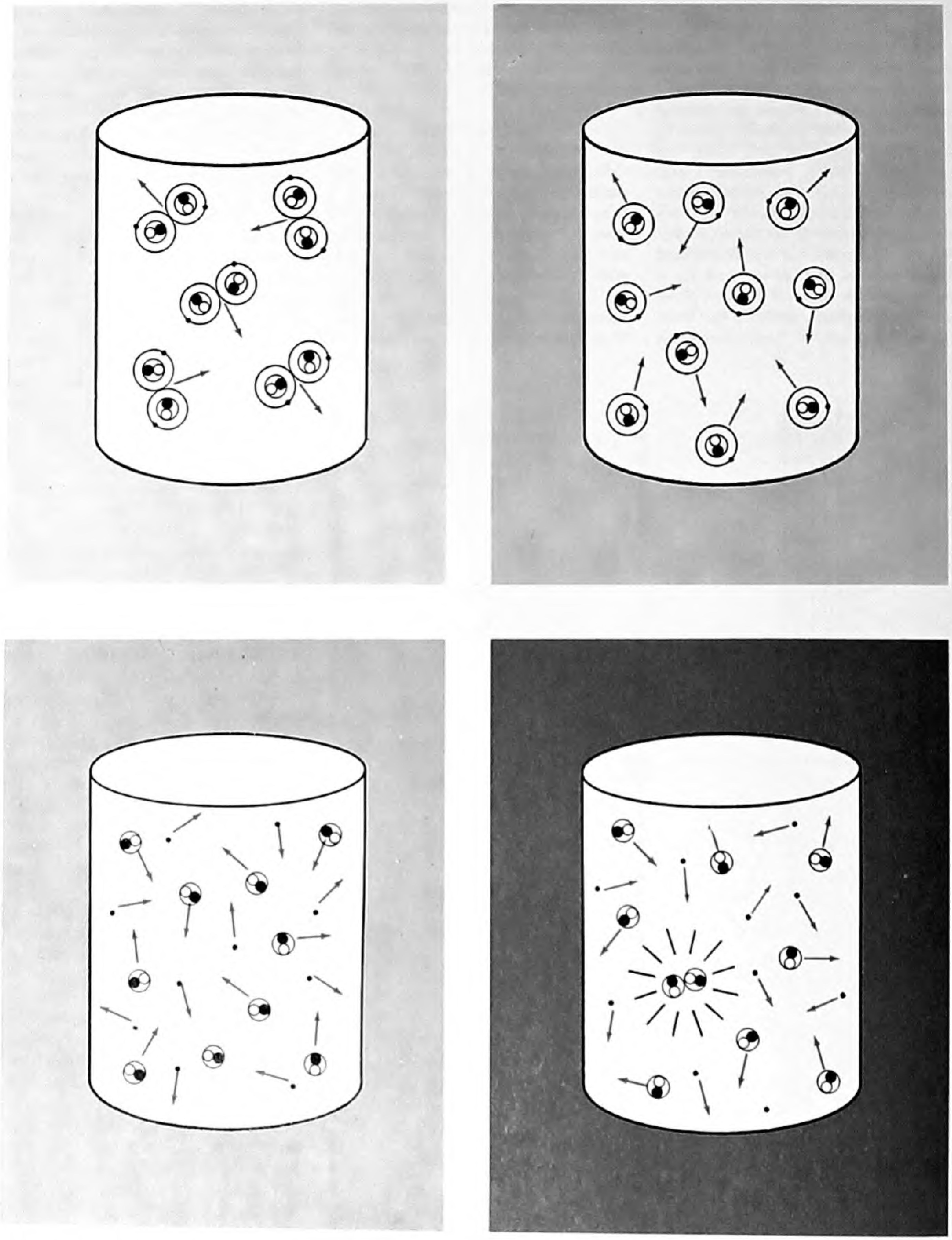
Theoretical calculations showed that it would take a very large current indeed—millions of amperes—to confine a plasma of high temperature and low density. Not discouraged by this fact, investigators in many countries carried out experiments with simple pinch tubes [*see illustration on page 294*]. They applied high voltage to a low-pressure gas in an insulated tube and produced an electrical discharge. This ionized the gas, and heavy current then began to flow. As they hoped, the pinch made its dramatic appearance. But with it also came a blow to their hopes. The pinch lasted only a millionth of a second or so; no sooner had the column of plasma been compressed than it writhed violently and drove itself to the tube wall. Furthermore, the tighter the pinch, the faster it destroyed itself.

This was not hard to understand, and in fact was predictable theoretically. Two different types of instability can develop. In the first place, any small kink in the pinched column will grow rapidly, because the magnetic pressure is stronger on the concave side of the kink (where the lines of force are crowded together) than on the convex side [*see first diagram at left below*]. The second cause of instability is a kind of "sausage" effect [*middle diagram at left*]. The plasma tends to pinch or neck itself off at one or more points along the column, and thus cuts itself into pieces.

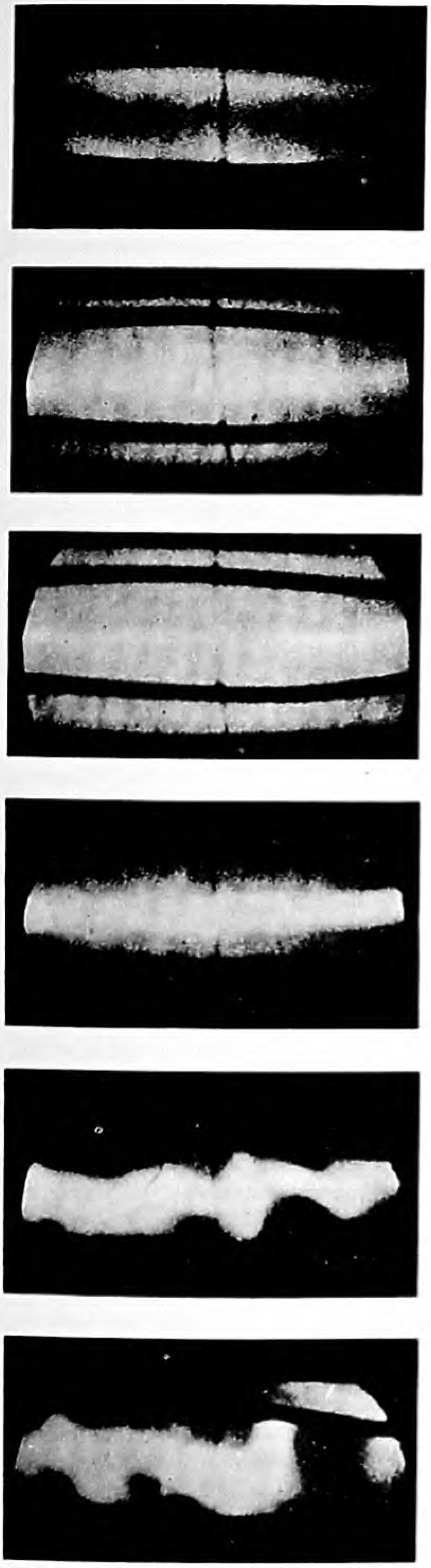
Incidentally, in connection with the latter phenomenon there is an interesting story which illustrates how hopes can suddenly rise and just as suddenly fall in an important but uncharted field of research such as the fusion power enterprise. When investigators first produced strong pinches in deuterium gas, they were delighted to discover bursts of neutrons—evidence of fusion reactions in the plasma. They thought they had reached thermonuclear temperatures momentarily. But on analysis they had to conclude that it was merely some obscure electrical effect, associated with the violent disruption of the pinch by the sausage instability, that had accelerated a few deuterons to fuse.



through a plasma. Tension would straighten kinks (*left*) and mutual repulsion would resist constrictions (*center*). Circular lines hold plasma away from conducting wall (*right*).



IMAGINARY EXPERIMENT shows the results of heating deuterium gas. At top left the gas is at room temperature and atmospheric pressure of 15 pounds per square inch. Heated to 5,000 degrees centigrade (*top right*), the diatomic molecules have split to form deuterium atoms and the pressure is 600 pounds per square inch. When the temperature is raised to 100,000 degrees (*bottom left*), atoms are ionized, forming a plasma of electrons and deuterons. The pressure is 20,000 pounds. At a temperature of 100 million degrees (and a pressure of 22 million pounds) some deuterons can fuse, releasing energy, but the reaction is still not self-sustaining.



LIFE OF A PINCH is depicted in this series of image-amplifier photographs. The top three pictures show the formation of the pinched column of plasma (*bright band*) as current is sent through the gas. The remaining views show its death, brought on by the development of kink instabilities.

How could the pinch be stabilized? Theoretical investigations published in the U. S., in the United Kingdom and in the U.S.S.R. have suggested a possible answer, although they have not yet shown how the physical conditions necessary to make it work could be achieved. The idea is to create not only a pinching magnetic field around the plasma but also a strong longitudinal magnetic field *within* the plasma column [*see diagrams on pages 300 and 301*]. The internal field would act as a stiffener. If a kink started to develop, it would tend to stretch the interior lines of force, and their elastic resistance would pull out the kink. Similarly if the sausage type of constriction tried to pinch into the column, the internal lines of force would resist being squeezed together and thus would prevent collapse of the column.

There is a third type of instability which could destroy a plasma column: namely, a long, gentle bend of the column that would grow in strength and push the column to the chamber wall [*see diagram at right on page 301*]. However, this can be counteracted by using a conducting material for the walls of the tube. Since a conductor acts as a barrier to a magnetic field, the magnetic field lines around the plasma column would be crowded against the wall where the bent column approached it, and the resulting back pressure would push the column back toward the center of the tube.

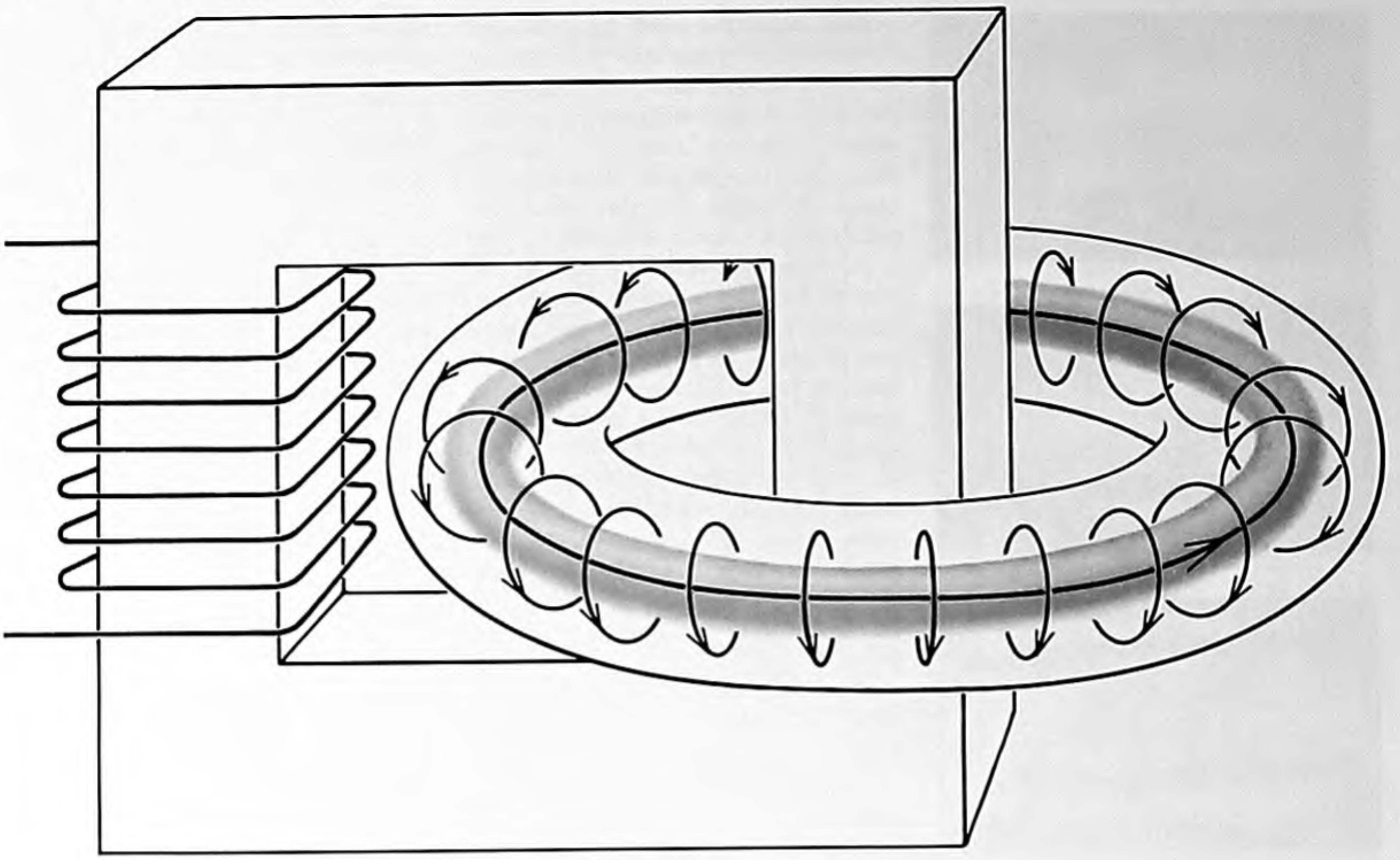
I should make clear that the straight pinch columns illustrated here are simplified systems which merely exemplify the principles. In practice it would probably not be desirable to try to produce a stable pinch in a straight tube, for several reasons: among other things, the electrodes at the ends of the tube would have a cooling (*i.e.*, quenching) effect on the plasma. Pinch experiments have already been performed with other shapes. One of these is a doughnut-shaped tube in which currents are induced and can circulate without bumping into a solid surface. A high voltage applied to the winding around a large iron transformer core in the tube produces an electrical discharge in the gas, which then functions as a one-turn secondary winding. Very heavy currents can be induced into the plasma in this way.

Besides the pinch effect, other possible methods of forming a magnetic bottle are being investigated, in the U. S. and no doubt in other countries. Most of the current research on the fusion power problem involves experiments in very complicated electromagnetic phenomena—the behavior of a high-temperature plasma in various magnetic field configurations. It is extremely hard work, taxing all the resources of our present technology. As I have mentioned, there is little likelihood of anything like a sudden "break-through" to a radically simpler approach. Since we shall make headway only by many mutually reinforcing studies and by the development of new and unusual techniques, it is evident that progress would be more rapid if the various nations involved joined in a cooperative exchange of information on all phases of the research.

Tracing the Wisp

We are having to learn for the first time how to handle and manipulate rapidly huge quantities of electrical energy, and to build big, extremely clean vacuum systems. One of the thorniest problems is measurement. It is hard enough to create a hot plasma in the first place, but to find out what it is doing, once created, is sometimes still harder. Some new measurement techniques have already been developed. To measure the density of a plasma there is a new "microwave interferometer" using beams of millimeter radio waves as probes. The temperature of the plasma is assessed by studying its X-ray emissions, its radio "noise" and its escaping particles or reaction products. Somewhat surprisingly, a very hot plasma emits little visible light. But it has been possible to make spectroscopic studies of cooler plasmas, in which the atoms are not completely stripped of their electrons. These measurements not only indicate the temperature but also tell the velocity of the plasma's motions, by the Doppler shift. And simply by measuring changes of the magnetic field during an experiment we can get information about the temperature, density, shape and velocity of plasmas as they are formed.

To appreciate the observation problem you have to picture the scene of the events we are studying. Most of the experiments on high-temperature plasmas are carried out within antiseptic vacuum systems surrounded by conductors carrying large and rapidly varying electrical currents. And what we have to detect are the whims of an invisible, short-lived wisp of near-nothingness somewhere in the bowels of this apparatus!



DOUGHNUT-SHAPED PINCH TUBE is made by threading a transformer core through a hollow ring containing plasma. Current through the winding at left causes a strong induced current in the plasma, which is then pinched by its own circular magnetic field.

Tapping the Power

Plainly it will be several years before any fusion reactor is developed to the point where we have to face the final problem of extracting its power. But of course some thought has already been given to this matter.

If the fuel is a mixture of deuterium and tritium, the lion's share (80 per cent) of the energy released by the fusion is carried off by the neutron emerging from the reaction. This energy would be tapped by trapping the fast neutrons and feeding the resulting heat to a steam system generating electricity in the conventional way. Since such a reactor would have to breed more tritium, the neutrons would probably be trapped in a breeding blanket of lithium surrounding the reactor. Some of the electrical power generated would have to be fed back to the reactor to maintain the magnetic bottle.

If the fuel is deuterium alone, we have the intriguing possibility of turning the energy output directly into electricity. In the deuteron-deuteron fusion, 66 per cent of the energy released is imparted to the charged reaction products—helium nuclei and protons. It is quite possible that conditions could be arranged so that most of these particles stayed trapped within the plasma. In that case the heated and expanding plasma would tend to push outward against the magnetic field, and by the use of properly arranged circuits this motion could in principle be made to generate a current. In other words, in pushing against the magnetic field the expanding plasma would do work, just as steam expanding against the piston of a steam engine does work. In the case of the plasma the piston would be the "wall" of the magnetic field, and the linkages converting its energy into useful work would be electrical circuits instead of rods and wheels. It is possible that the efficiency of this engine might be much higher than that of the conventional steam cycle, for the thermodynamic principles that limit the efficiency of ordinary heat engines would not apply.

In the last analysis the feasibility of fusion power probably will hinge primarily on the size of the reactor. Very likely many potentially workable reactor schemes will have to be rejected on the basis that they would be impracticably large. On the other hand, "pocket-edition" fusion reactors are simply out of the question, on theoretical grounds. The fusion power plants of the future, if they are ever realized, will in all likelihood be large central stations for generating electrical power.

Growing Hopes

The scientists working on the fusion power project in the U. S. confidently believe that all the problems will eventually be solved, for, difficult though the problems are, they now see no really fundamental barrier standing in the way of ultimate success.

I want to mention a few of the milestones that have been passed, insofar as secrecy restrictions permit. The first group to consider the possibility of obtaining fusion power through the magnetic confinement of a hot plasma, so far as we know, were Enrico Fermi, Edward Teller, James Tuck and others at the Los Alamos Scientific Laboratory. Although they advanced their ideas around the end of World War II, no extensive experimental work was started in the U. S. until about 1951, when

THE "PERHAPSATRON," an instrument for producing the pinch effect in a plasma, was photographed at the Los Alamos Scientific Laboratory. Doughnut tube is in the center.

programs began under Tuck at Los Alamos and Lyman Spitzer at Princeton. Scientists in the United Kingdom apparently had launched some work two or three years earlier, and the U.S.S.R. may have started about the same time. In 1952 the U. S. Atomic Energy Commission sponsored a large conference on controlled-fusion reactions. In the same year a program was initiated by Herbert York at the University of California Radiation Laboratory in Livermore. Smaller programs have since been set up at the Oak Ridge National Laboratory and New York University. Within the past year two substantial programs have been initiated with private capital: at the General Atomics Laboratory in San Diego, Calif., and at the General Electric Research Laboratory in Schenectady, N. Y.

Among other countries known to be engaged in studies of controlled fusion are France, Germany, the Netherlands and Sweden. It is abundantly evident that the search for fusion power is taken seriously in all parts of the world. The nature of the problems being faced, the time it will no doubt take to solve them, and the importance of the goal to be won are compelling reasons to hope for the growth of international cooperation in the research.

If the reader of this article is left with the impression that the search for fusion power is at once the most fascinating, the most difficult and potentially the most important peacetime scientific effort ever undertaken, he will be sharing the opinion of the many scientists now working on this problem.

The Author

RICHARD F. POST is a physicist at the Radiation Laboratory of the University of California. Born in Pomona, Calif., he attended Pomona College. During World War II he served in the Naval Research Laboratory, where he won a Meritorious Civilian Service Award for his work on underwater sound. In 1950 he received a Ph.D. from Stanford University. He has been connected with the University of California since 1951.

Bibliography

CONTROLLED FUSION RESEARCH: AN APPLICATION OF THE PHYSICS OF HIGH TEMPERATURE PLASMAS. Richard F. Post in *Reviews of Modern Physics*, Vol. 28, No. 3, pages 338-362; July, 1956.

SCIENTIFIC AMERICAN October 1954, Vol. 191, No. 4, pp. 62-66 OFFPRINT 237

PHOSPHORS

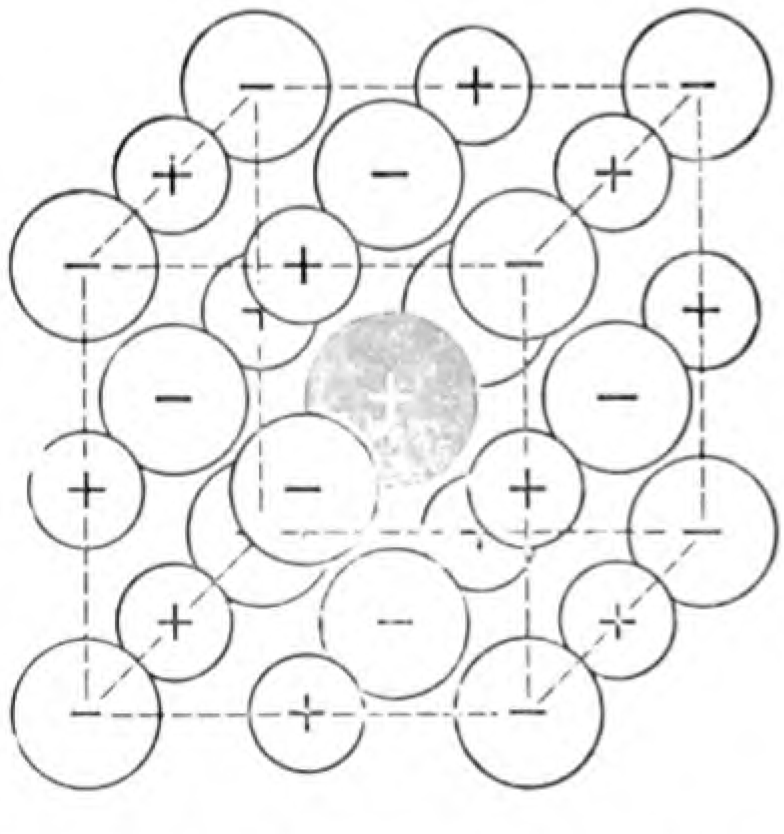
by J. S. Prener and D. B. Sullenger

In their studies of these luminescent materials physicists have reached the point where they can formulate verifiable theories about how they function and have begun to make better phosphors than nature.

These blazes . . . giving more light than heat . . .
Hamlet, Act I, Scene 3

Although Polonius in this remark to his daughter Ophelia had another subject in mind, he might have been speaking of luminescence, for luminescence is certainly an emission more of light than of heat. This production of "cold light" has long been a source of amazement. In early times men regarded luminescent animals and minerals as possessing magical qualities. In 1603 a Bologna shoemaker, who apparently practiced alchemy on the side, heated together a mixture of charcoal and barites and accidentally produced the first recorded synthetic phosphor. Instead of the philosophers' stone, he obtained a material which glowed in the dark after exposure to sunlight.

We employ today a multitude of organic and inorganic phosphors—in fluorescent lighting, in television, in radar, in paints and inks, in instruments for detecting various radiations. The fluorescent lamp, first exhibited to the U. S. public at the 1939 New York World's Fair, now challenges the incandescent lamp in popularity. The exploitation of the properties of phosphors has removed them from the realm of the laboratory curiosity. Some phosphors are now being produced in tonnage quantities, though most of their uses require only gram amounts in each device. The total annual world production of phosphors is estimated to be more than one million pounds.



LATTICE of potassium (+) and chloride (—) ions surrounds thallium ion (*shaded*).

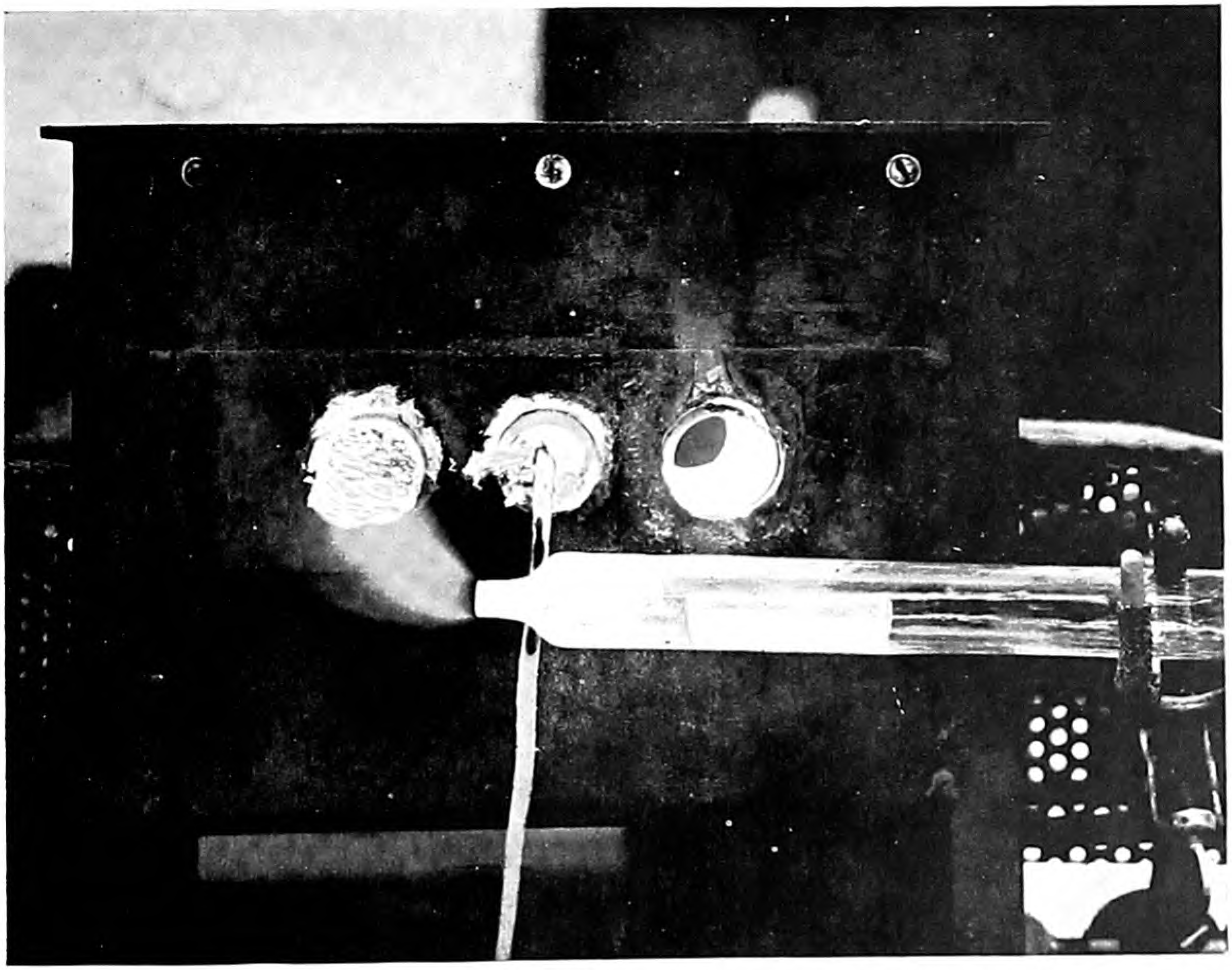
The word "phosphor" comes from a Greek word meaning light-carrier. From the same word we get the name of the element phosphorus, which, upon exposure to natural weathering, emits light. This has been shown to be due to the chemical release of energy by oxidation of the element. Photoluminescent phosphors are substances which absorb electromagnetic energy, usually ultraviolet light, and then re-emit this energy in the form of visible light. The absorption of energy by the phosphor is called "excitation" of the phosphor, and the release of the energy is called fluorescence. The difference in the wavelength of the absorbed and emitted light is an important property of phosphors, because the material as a consequence does not reabsorb much of the light it emits.

The photoluminescent phosphors are not easy to prepare, for their composition must be rigidly controlled. Because of this, and because of the desirability of having exact knowledge as to their physico-chemical properties, much research has centered upon them in the past two decades.

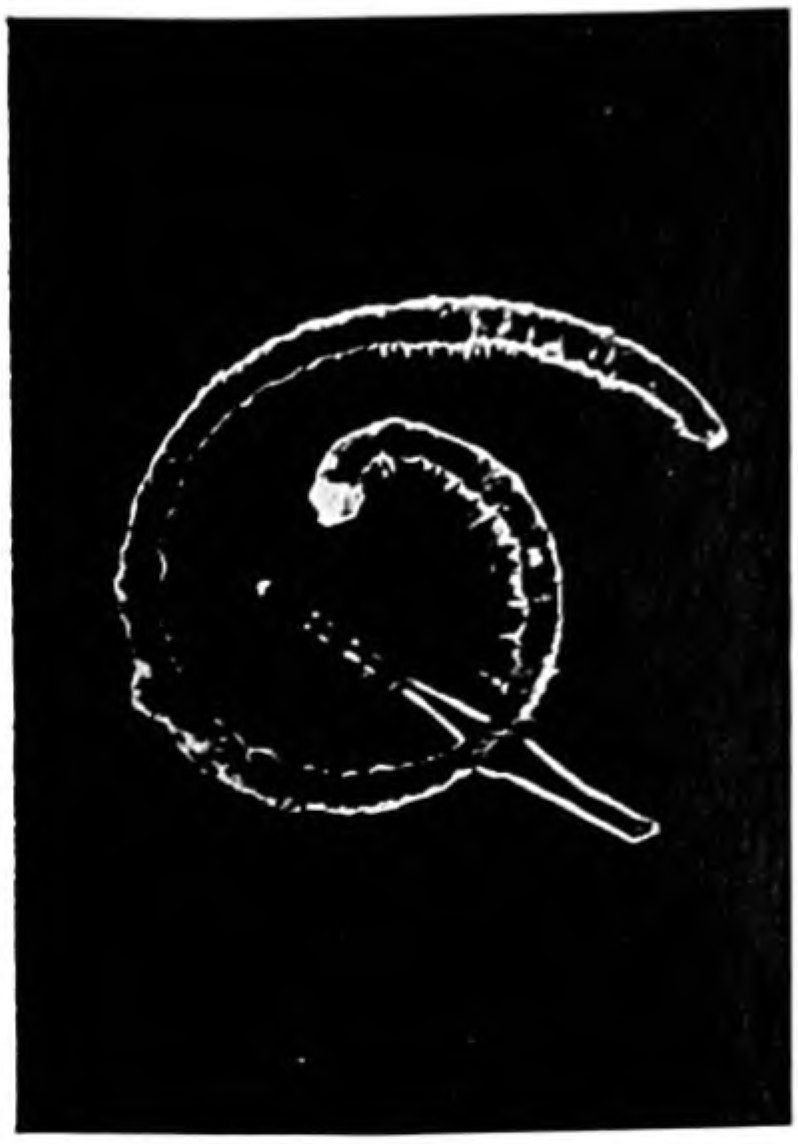
In the fluorescent lamp the phosphor is coated on the inside of the glass tube, and the exciting ultraviolet light comes from a mercury glow discharge. There are many natural phosphors; typical ones are the minerals willemite (zinc orthosilicate), wurtzite (zinc sulfide) and fluorite (calcium fluoride). The phosphor chemist has been able to synthesize, under carefully controlled conditions, much more efficient phosphors than those found in nature. The property of fluorescence is known to be due to the presence of small amounts of impurities, called activators, in the compound. The specific properties of a phosphor (*e.g.*, the color of the light it emits, the wavelength of light it must absorb to be excited to fluorescence, the brightness of the emitted light, the duration of phosphorescence, or afterglow) depend primarily on the chemical nature of the material and the activator it contains. In the zinc orthosilicate phosphor (green fluorescence) the activator is manganese; in zinc sulfide (blue fluorescence) it is silver; in calcium fluochlorophosphate (white fluorescence) it is antimony and manganese. The needed impurities amount to less than 1 per cent.

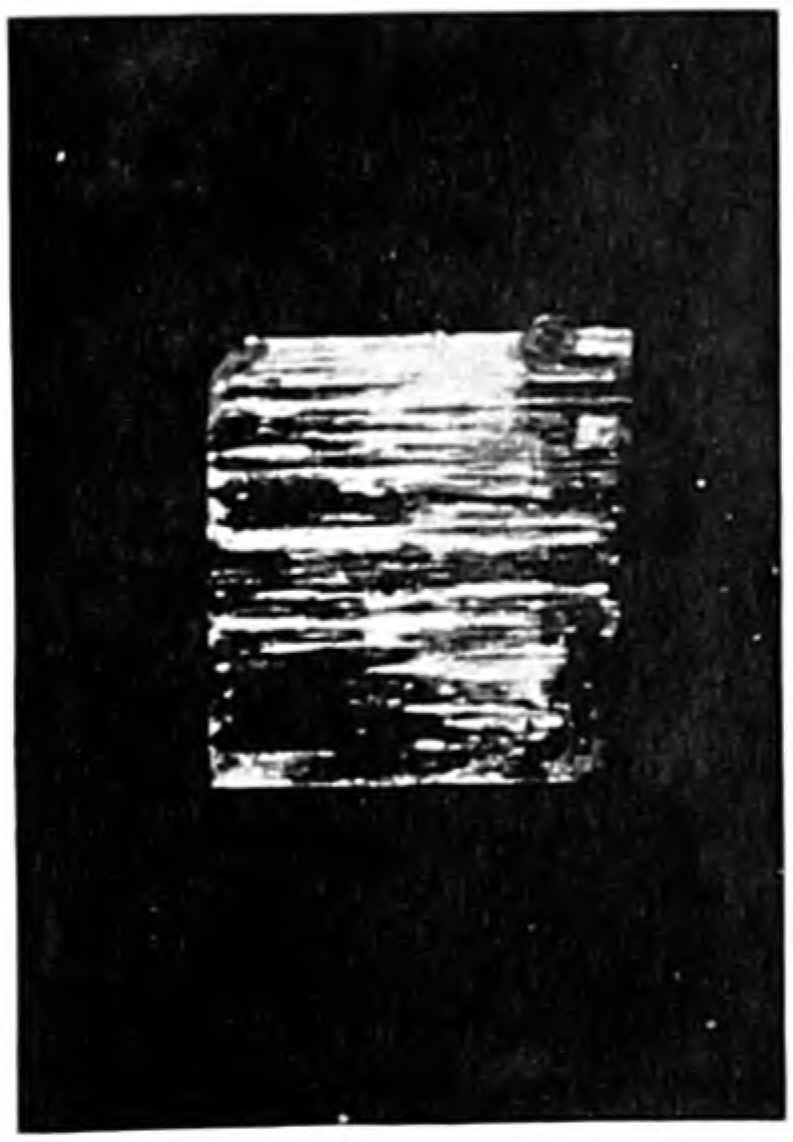
Let us see how a chemist prepares a typical phosphor: manganese-activated zinc orthosilicate. He takes very pure zinc oxide and silicon dioxide, adds .1 per cent manganese in the form of a salt such as manganese sulfate, mixes these ingredients together and heats them to 1,100 degrees centigrade. At this temperature the zinc oxide and silicon dioxide react to form zinc orthosilicate (Zn_2SiO_4) and the manganese atoms diffuse into the crystals. Many other phosphors are made as polycrystalline powders by similar solid-state diffusion methods. Some phosphors can also be prepared as single crystals and as transparent thin films about one micron thick.

A great deal of research and development has gone into the improvement of old phosphors, synthesis of new ones

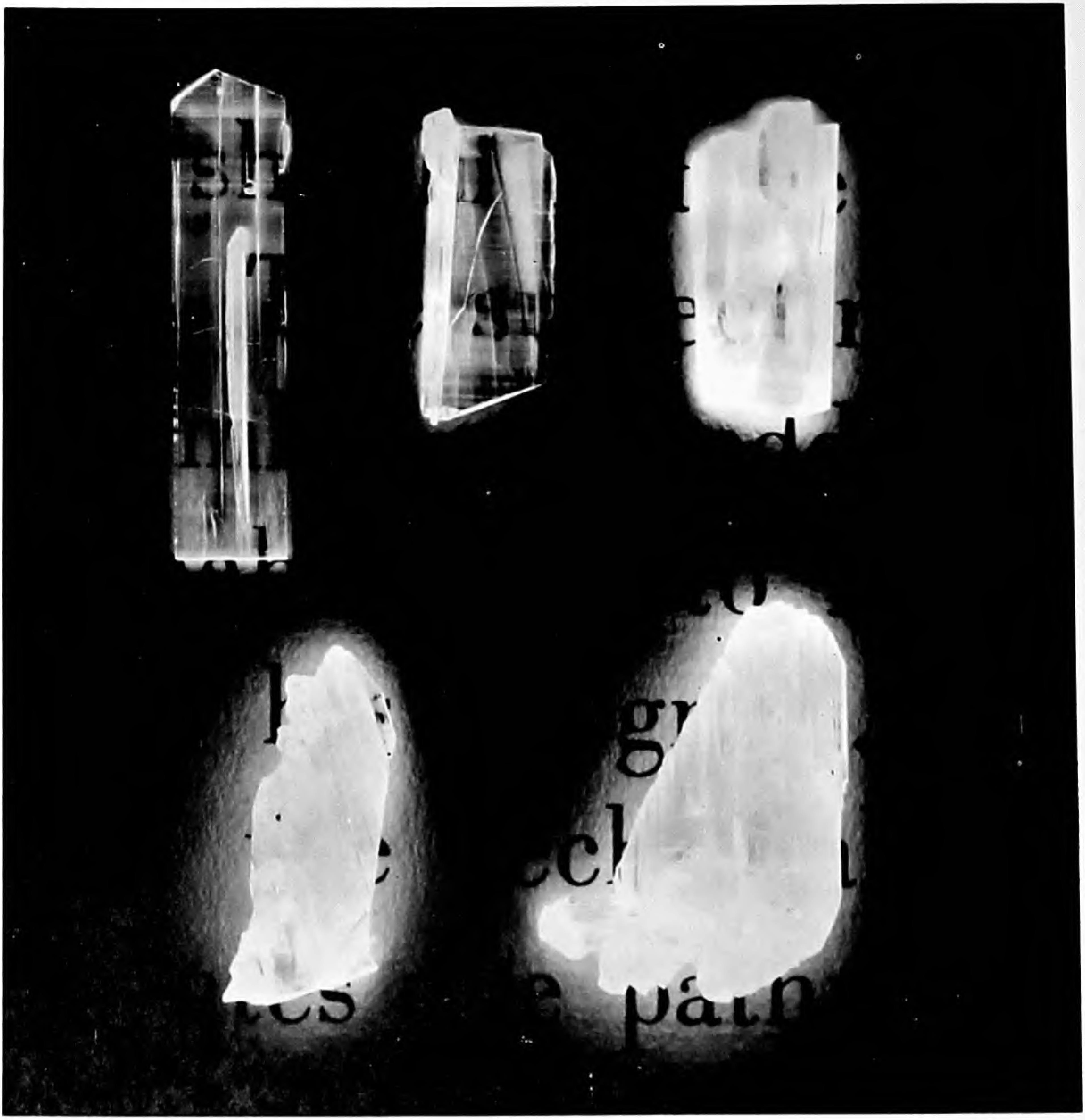


PREPARATION of experimental zinc sulfide phosphor in the General Electric Research Laboratory is accomplished by mixing powdered ingredients in a quartz boat and firing them in an electric furnace at 1,000 degrees in atmosphere of hydrogen sulfide.





SINGLE CRYSTALS are shown photographed by an external light. At left is a spiral form of zinc sulfide; center is cadmium sulfide; at right, potassium chloride activated with manganese. They are not in scale; the first two are five millimeters wide; the third, 20.



PHOSPHORESCENCE of six pure crystals of zinc sulfide is evident in this photograph taken by light emitted from them during excitation from an ultraviolet light source. Zinc sulfide occurs in nature as the mineral wurzite, and it emits a bluish light when it fluoresces.

and attempts to understand the physics of fluorescence. Most of the development of new phosphors and the improvement of old ones has proceeded along purely empirical lines. Recently, however, detailed studies of the physics of luminescence have enabled us to understand the phenomenon well enough to begin designing phosphors for specific applications from theoretical considerations. For instance, on the basis of studies of one simple phosphor (thallium-activated potassium chloride) it was predicted that use of mercury as the impurity in a material such as the mineral harmotome, a silicate of aluminum, barium and potassium, would yield a phosphor capable of absorbing light at 2,537 Angstroms and emitting blue light. Such a phosphor has already been prepared. As more complicated phosphors are treated theoretically, no doubt more such analogous cases will arise.

In predicting the luminescent properties of a phosphor, the theory of luminescence starts with the properties of the ions that make up the material and the known forces of interaction between them. The only phosphors studied theoretically so far have been those that are the simplest to treat mathematically. The first was the one mentioned above—thallium-activated potassium chloride.

This phosphor consists of potassium chloride with monovalent thallium ions randomly distributed in the crystal, replacing about .1 per cent of the potassium ions. Let us look at one little region

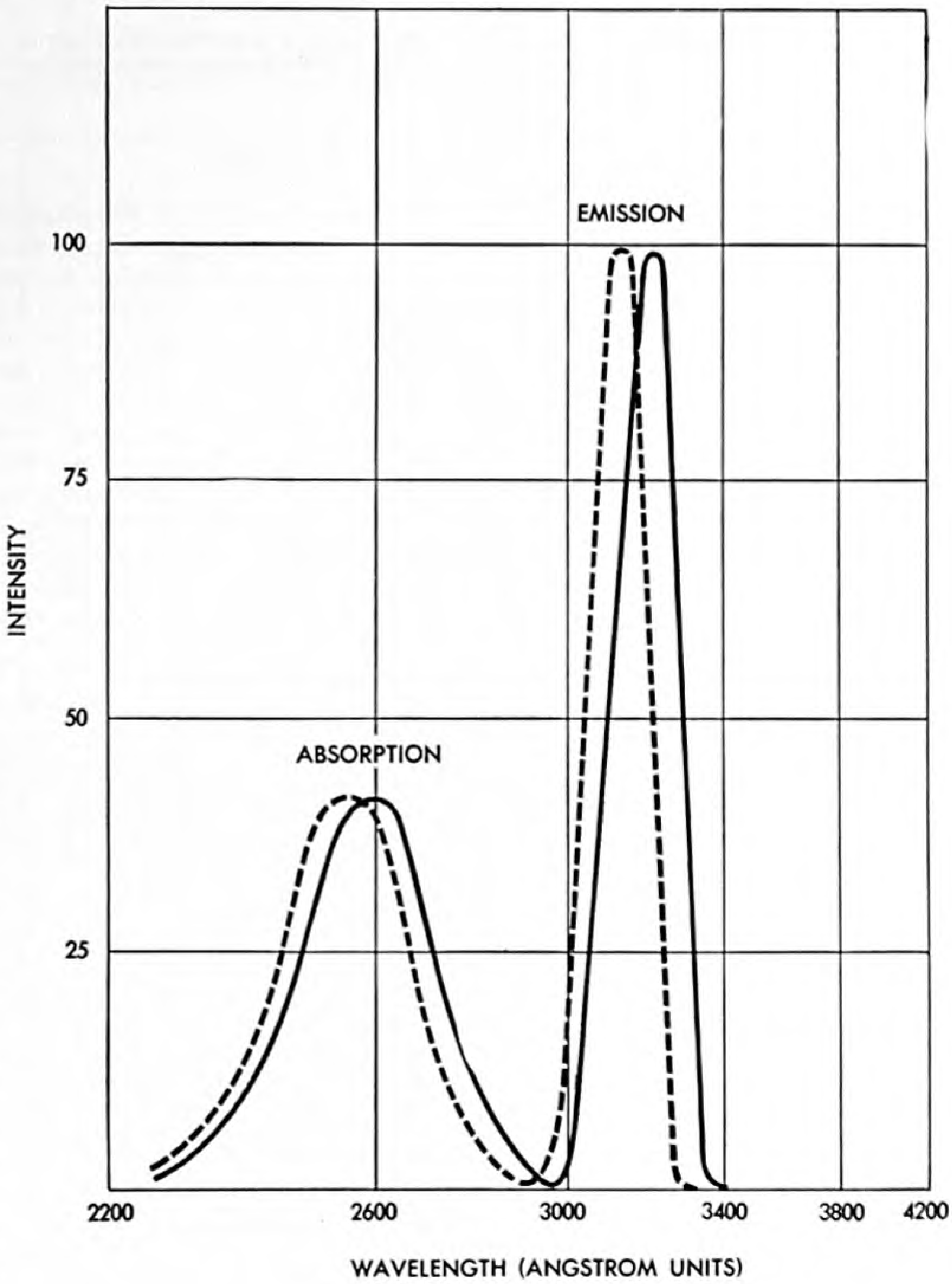
of the crystal in the vicinity of a typical thallium ion [*see diagram on page 306*]. The thallium ion is closely surrounded by six chloride ions, and farther away are other chloride and potassium ions. The whole crystal can, of course, be built up by a repetition of this fundamental building block in three dimensions, except that a potassium ion usually sits where the thallium ion is in this diagram.

What holds the crystal together? We first recognize that potassium chloride, being an ionic solid, is built up of positively charged potassium ions and negatively charged chloride ions, which are attracted to one another by virtue of the electrostatic forces between charged particles. When two ions get too close together, however, the overlap of the electron clouds surrounding their nuclei results in repulsion. This force is not due entirely to the electrostatic repulsion between the negatively charged electrons; it can be understood in detail only on the basis of quantum mechanics. At all events, the equilibrium distance between ions in a crystal is regulated by a balance of the attractive and repulsive forces. It takes energy to pull the six chloride ions farther than this distance from the thallium ion or to push them closer to it. Of course the ions in the lattice are not stationary; they vibrate around their equilibrium position.

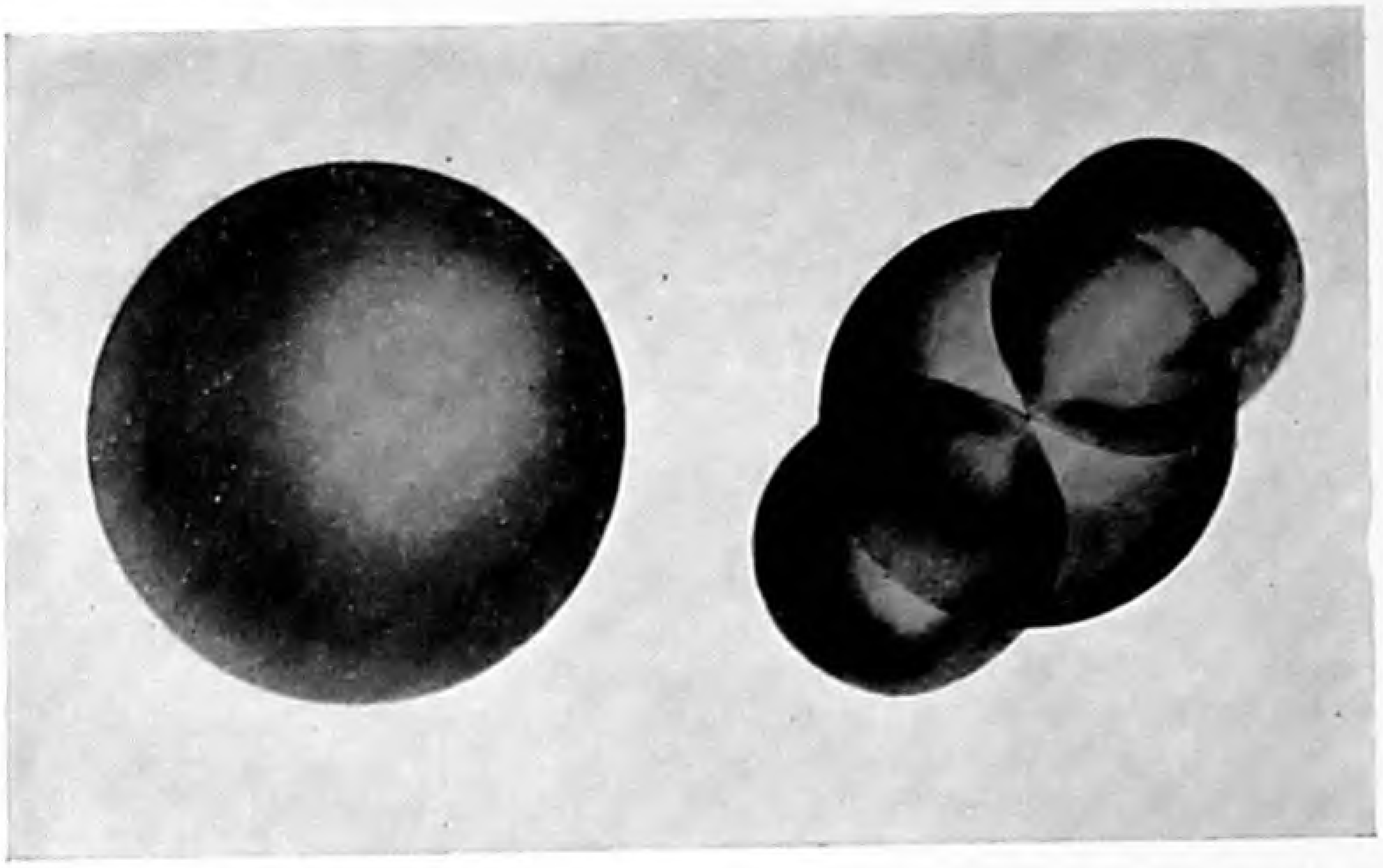
The thallium ion and its six surrounding chloride ions are called the activator system. When the six chlorides are at any given distance from the thallium, we have what is called a "configuration" of the system; the lowest-energy configuration is the one at the equilibrium distance, and those corresponding to higher energies are said to have positive configuration coordinates. The configuration coordinate is the distance between the thallium ion and the surrounding chloride ions. We can show the energy of the activator system graphically as a function of the configuration coordinate if we can calculate in detail what energy is required to move the six chloride ions from their equilibrium position to some new position. This energy can be calculated by the methods of quantum mechanics. It depends on the specific distribution of the electrons around the thallium ion and around the chloride ions, for the repulsive forces are particularly sensitive to these distributions. We would like to re-emphasize that the activator system takes up these various configurations because of the thermal vibrations of the chloride ions around their equilibrium positions.

Now let us see what happens when we shine ultraviolet light on this thallium-activated phosphor. It can absorb only a certain range of wavelengths. Each thallium ion will absorb a photon of this ultraviolet radiation. This absorption will increase the energy of the system by a certain amount depending upon the wavelength of the radiation. In the higher energy state the electrons are rearranged in a different way around the nucleus of the thallium ion. This distribution can again be calculated by the methods of quantum mechanics. Because of the new distribution, the repulsive forces between the thallium ion and the six surrounding chloride ions are less than they were in the unexcited thallium ion. Hence the equilibrium position of the chloride ions is different: they are closer to the thallium ion than in the ground state. The difference in the shapes of the electron distributions in the unexcited and excited states of the thallium ion is shown schematically in the model on the next page.

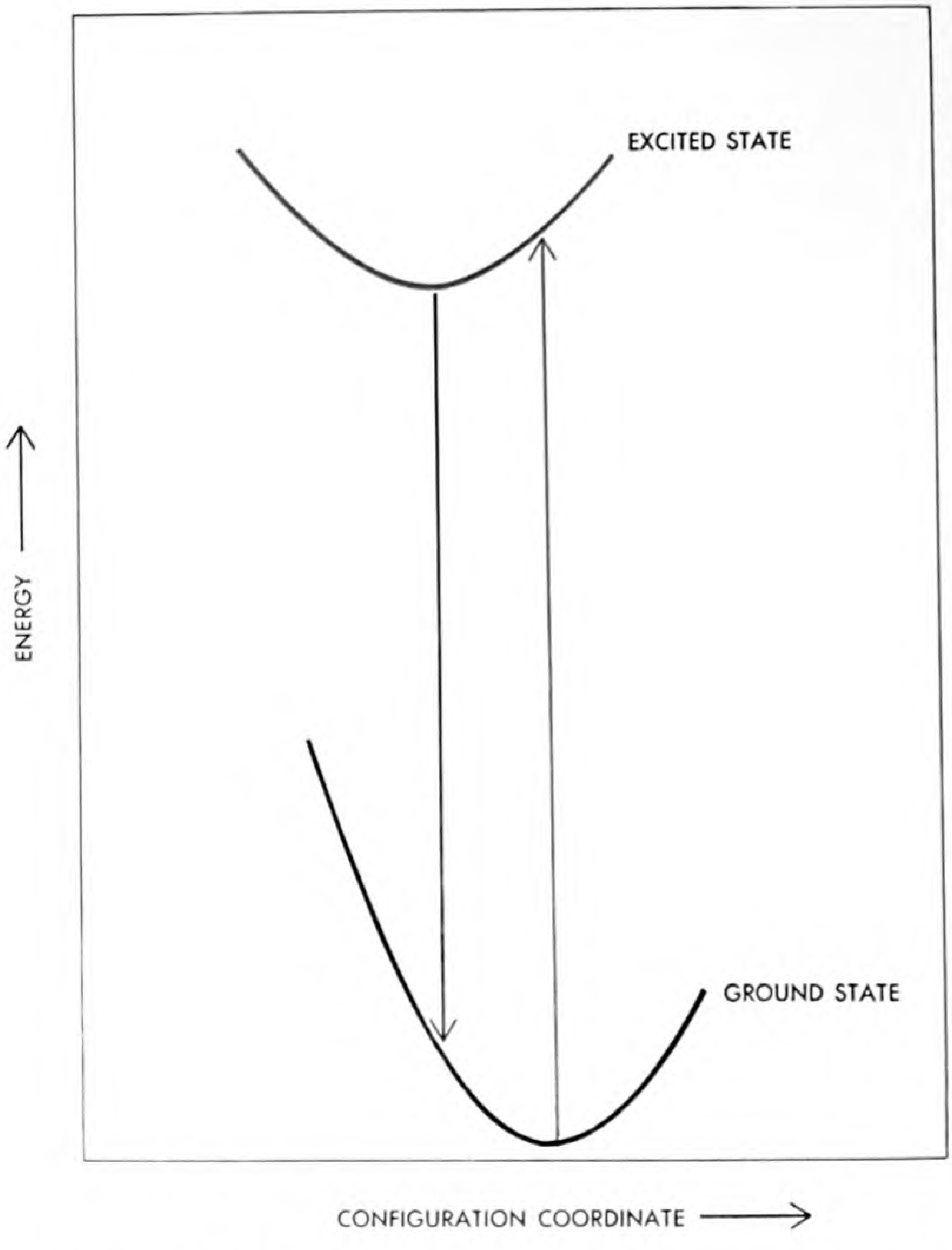
From curves representing the energy configurations in the two states we can derive theoretically the absorption and emission spectra of the phosphor [*see chart on next page*]. The peak of the absorption spectrum will correspond to the energy difference shown by the arrow at the right on the chart, and the peak of the emission spectrum to the arrow on



ABSORPTION AND EMISSION curve theoretically calculated (*solid line*) for thallium activated potassium chloride agrees well with the experimental one (*broken line*). The peak of absorption of light comes at shorter wavelength than peak of emission.



ELECTRON CLOUD of a thallium ion differs in the unexcited state (*left*) from that in excited state (*right*). The model shows difference in electron distributions around nucleus.



ACTIVATOR SYSTEM of the thallium ion is shown by the graph. As it absorbs one photon of ultraviolet light, its energy is raised by the amount of the photon's 5.48 electron volts from the ground state (*lower curve*) to the excited state (*higher curve*). The system comes to equilibrium. Then ion emits one photon of light, returning to the ground state.

the left. The difference in length of the two arrows indicates that the peaks in the absorption spectrum and in the emission spectrum come at different wavelengths. It is this property that makes the phosphor transparent to its own emission, as we have mentioned. The exact shape of the absorption and emission spectra can be calculated from a knowledge of the relative probability of various configurations. The curves of these theoretical spectra come very close to those based on actual observation [*chart on page 309*].

We can sum up the absorption and emission processes in a phosphor briefly as follows: The most probable configuration of the activator system is the one of lowest energy. When we shine ultraviolet light on the phosphor, the thallium ion absorbs a photon of ultraviolet light and is excited to a higher energy state. The energy rise is represented by the length of the arrow at the right in the chart. The activator system adjusts itself to a new equilibrium position, corresponding to the upper end of the arrow at the left in the chart. The difference in energy between this level and that of the tip of the arrow at the right appears as heat, which is dissipated in the crystal. The excited thallium ion now emits a photon of visible light which appears as luminescence, and the energy of this photon is equal to the length of the arrow at the left. The activator system now readjusts itself to its equilibrium configuration in the unexcited state, and again the difference in energy appears as heat. Now the thallium ion is ready to absorb another photon of ultraviolet light and begin a new cycle. It takes about a millionth of a second to complete one cycle.

Although this theory is only in its infancy, it has been applied quite successfully to a simple phosphor, predicting most of the important luminescent properties. The configuration coordinate model recently has been sustained independently by studies of the effects of pressure on the emission and absorption spectra. Furthermore, as we have noted, the theory has already yielded one new synthetic phosphor, and more complicated ones are under investigation.

The approach offers considerable hope of creating phosphors which will greatly improve the performance of devices utilizing luminescence and open up many new, large-scale applications. Fully as important as the applications is our increased understanding of the nature and properties of the solid state.

The Authors

JEROME S. PRENER and DON B. SULLENGER are phosphor chemists at the General Electric Research Laboratory and Cornell University, respectively. Prener, a graduate of Brooklyn College in 1944, worked as a research chemist in the atomic energy project from 1944 to 1946, then as a graduate student at the Polytechnic Institute of Brooklyn did research on phosphors under contract with the Navy. He joined General Electric in 1949. Sullenger, who graduated from the University of Colorado in 1950, tested petroleum and did hospital laboratory work for the U. S. Army during the Korean War. He is now at Cornell, as a research assistant in the department of chemistry.

Bibliography

AN INTRODUCTION TO THE LUMINESCENCE OF SOLIDS. Humboldt W. Leverenz. John Wiley & Sons, Inc., 1950.

FLUORESCENCE AND PHOSPHORESCENCE. Peter Pringsheim. Interscience Publishers, Inc., 1949.

LUMINESCENT MATERIALS. G. F. J. Garlick. Oxford University Press, 1949.

SCIENTIFIC AMERICAN May 1958, Vol. 198, No. 5, pp. 64-73 OFFPRINT 238

THE TEACHING OF ELEMENTARY MATHEMATICS

by E. P. Rosenbaum

Part II of a survey of the current ferment in science education. The outlook in mathematics: The next few years will bring substantial changes in high-school curricula.

If you have a youngster in high school and are occasionally called upon for help on his math, you may not remember the answers but you recognize the problems. It is the same old math that was taught a generation ago and indeed 200 years ago. The methods and the subject matter have not changed. But a revolution is impending. Deeply dissatisfied with the way mathematics has been taught, several influential groups of educators have begun to experiment with radically new methods of presenting the subject. There are high schools where sophomores are now taking home problems such as this: Prove that $[A, B/l/]$ if and only if ($A \notin l$ and $B \notin l$) and $\overline{AB} \cap l = \varnothing$. (Translated this reads: "Prove that a point B lies between another point A and a line l if and only if A is not a member of l and B is not a member of l and the segment AB does not intersect l.")

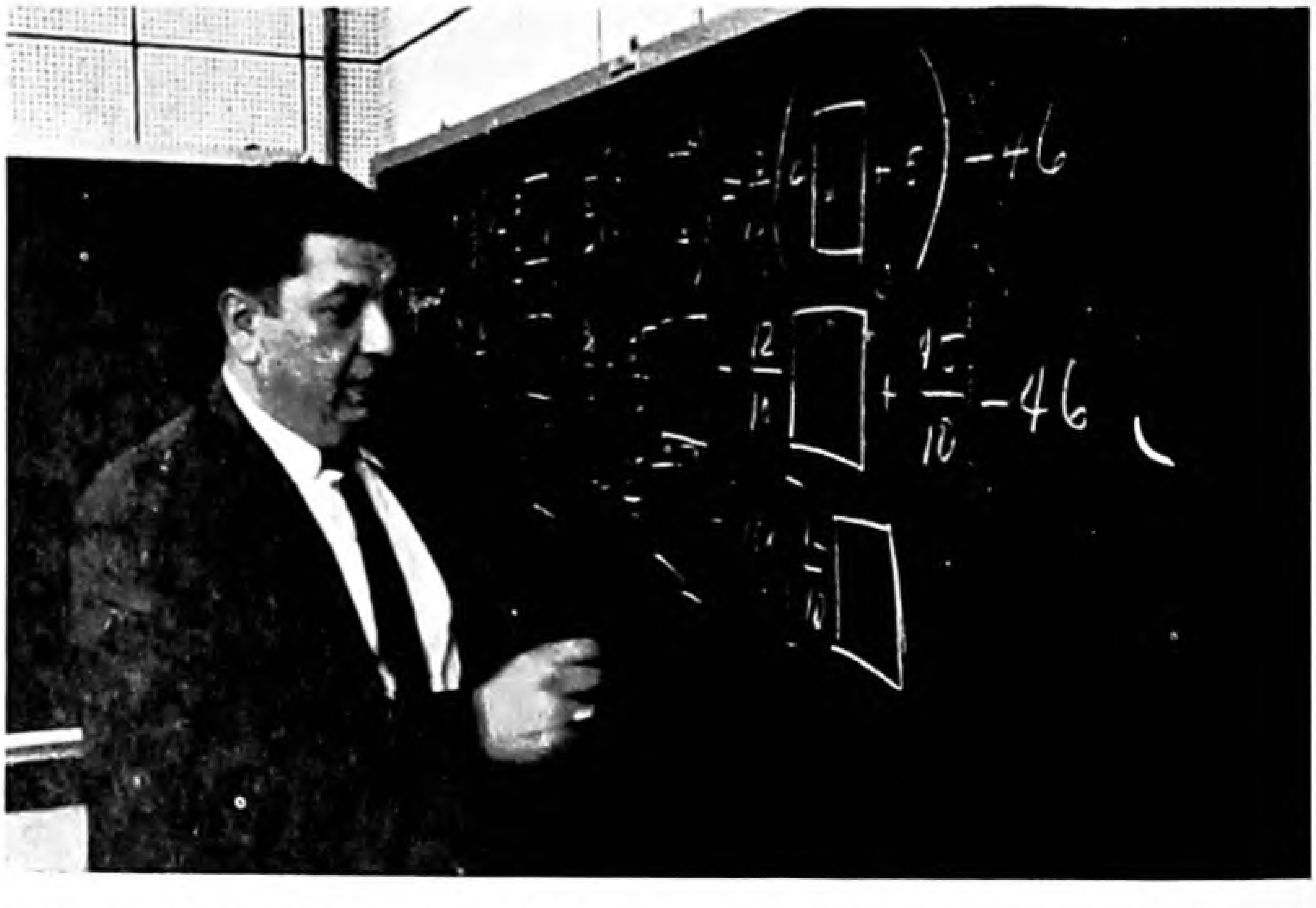
There are sharp differences of opinion about whether this sort of thing is helpful or will get very far. But on one thing almost everyone agrees: The old mathematics course must go. It does not tell the students what mathematics is all about. It does not give them any real understanding of the principles of the subject. It is so far behind the times that it leaves out practically all the new ideas and discoveries of the past 100 years. And above all, it has managed to make mathematics about the most unpopular of all branches of learning. Even cultivated men declare their ignorance of mathematics with a defiance akin to pride.

Something has to be done to make mathematics more meaningful and more exciting. At least three major attempts are under way. I shall try to report briefly the rationale and tactics of their approach.

Their underlying theme is to modernize the subject matter taught—more or less (some want to go modern farther than others). Going modern means mainly two things: pruning out dead wood and introducing some of the new fundamental ideas which within the last century have given more meaning and unity to all the traditional branches of mathematics.

As an example of dead wood one can cite the considerable attention devoted in trigonometry textbooks to solving triangles with logarithms. This was the only method available to surveyors and navigators a century ago; today technicians punch out the answer in calculating machines. Also in the category of dead wood are some of the classic Euclidean "proofs" of geometry; they are not really proofs at all. Today a mathematician attacks these problems with the calculus and the idea of limits, and arrives at truly rigorous proofs.

Mathematics as now taught in the lower schools and even in college seems to be a collection of separate subjects.



MODERN APPROACH to high-school mathematics is being tried by the University of Illinois Committee on School Mathematics. The photographs on these and the two following

Each has its own apparently arbitrary rules, taught by rote. But work on the foundations of mathematics in the last century has shown that all the branches of mathematics can be reduced to purely abstract terms, with common properties. As numbers are the elements of algebra, so points and lines are "primitive" elements of geometry, and we can deal with sets of either in the same way—indeed with the same operations. The rules of mathematical logic are universal. These basic ideas and logical processes can be taught even to school children, the modernists believe. In learning them, students will find mathematics more understandable and more meaningful. They will also get at least some acquaintance with modern thinking in mathematics.

The College Board Program

Foremost among the projects for modernizing the teaching of high-school mathematics is that of the Commission on Mathematics of the College Entrance Examination Board. This group, set up in 1955, is preparing a considerable revision of the high-school courses. Because the Commission represents a wide range of views, and because it hopes to exert an almost immediate influence on every school in the country, its program is comparatively conservative. It proposes to change the approach and spirit of the algebra course without altering the content substantially. It is revising the geometry course considerably. It wants to make certain changes in trigonometry. And it would add some entirely new courses to the high-school curriculum.

Algebra, like geometry, can be regarded as an abstract deductive system. It is built up from a set of undefined primitive notions and a number of assumed axioms. From these all the other rules and facts can be deduced by logical reasoning. The Commission does not propose to make high-school algebra purely an exercise in abstract deduction, but it believes that students should be led to appreciate the deductive nature of algebra, should learn what the axioms are and how they are used to prove some of the principles. The necessary skills in manipulating algebraic expressions will be easier to learn when the reasons behind the manipulations are understood.

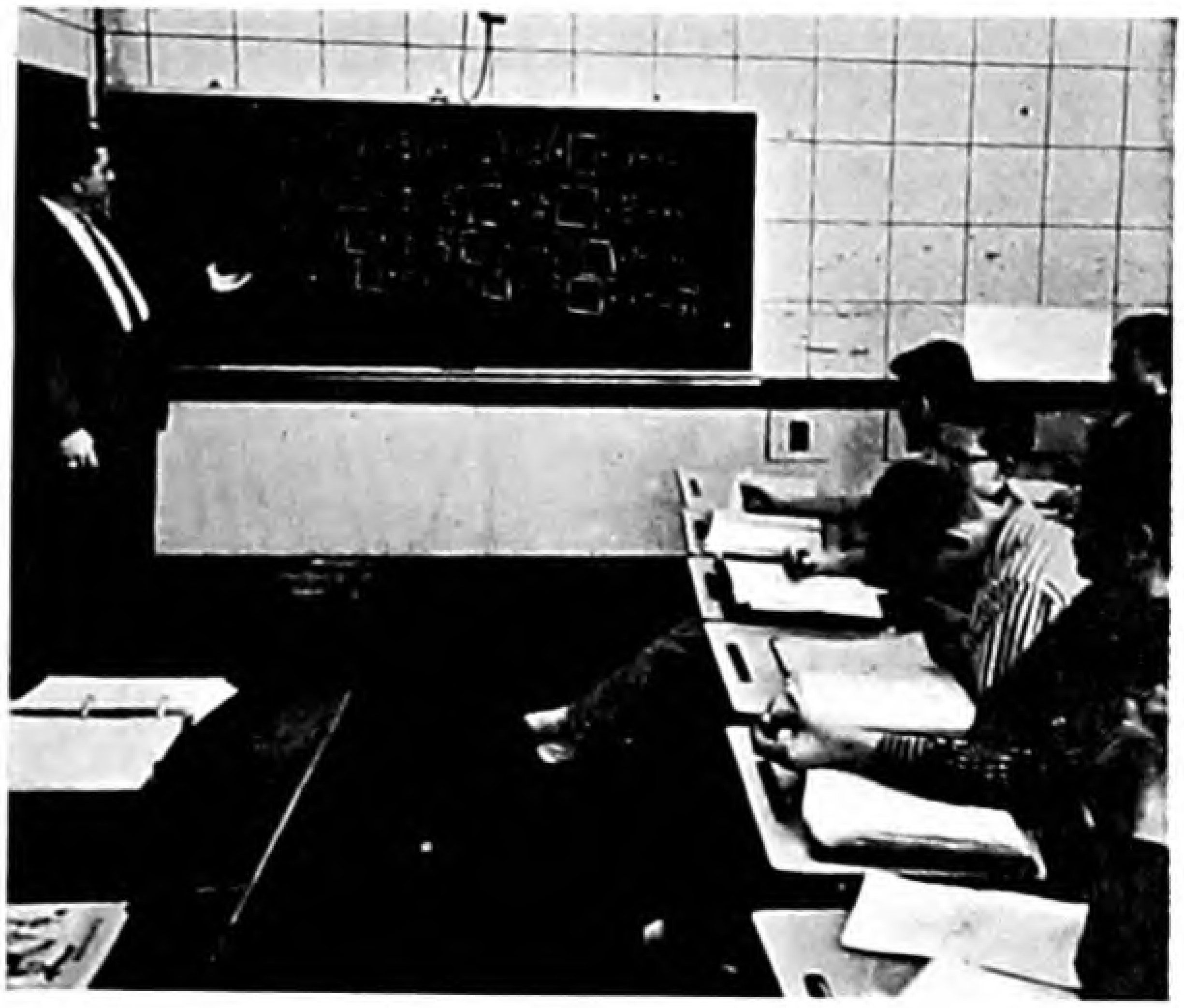
What are the "primitive" notions of algebra? They are the notions of number and of operations such as addition and multiplication. The axioms likewise are simple concepts—so simple that they hardly seem to need stating. If a and b are numbers, then $a + b$ is a number. If $a = b$ and $c = d$, then $a + c = b + d$ (if equals are added to equals, the sums are equal). Then there are the "commutative" laws, $a + b = b + a$, and $a \times b = b \times a$; the "associative" laws, *e.g.*, $(a + b) + c = a + (b + c)$, and the "distributive" law, $a (b + c) = ab + ac$.

Although they may appear mere platitudes, these rules form a mathematical system which underlies all the familiar manipulations of elementary algebra and provides a foundation for deriving further principles. Consider the theorem that the sum of two even numbers is also even. To begin the proof, evenness is defined as follows: A number n is even if and only if there is another number p such that $n = 2p$. The problem is: Given that a and b are even numbers, prove that $a + b$ is even. By the definition of evenness we can say that $a = 2x$ and $b = 2y$. By the axiom about adding equals, $a + b = 2x + 2y$. The distributive axiom says that $2x + 2y = 2 (x + y)$. Since x and y are numbers, $x + y$ is a number. Hence $2 (x + y)$ is an even number and so $a + b$ is even.

It may strike a layman that high-school freshmen will be neither attracted nor edified by such a laborious proof of a seemingly obvious idea. There are mathematicians who hold the same view. But the proponents argue that in a well-taught course exercises of this kind may become an intriguing journey into the realm of mathematical rigor.

The Theory of Sets

A central feature that distinguishes the "new" algebra is its use of the theory of sets, one of the most powerful tools of modern mathematics. A set is simply a group or collection. The books on a shelf, the people in a room, the letters of the alphabet, the numbers 1, 2, 3, 4 and 5—each of these groups is an example of a set. The set may have only one member or none at all (in which case it is the "empty" set). Or it may be infinite: *e.g.*, all the points inside a circle, or all the positive integers. Mathematicians commonly denote a set by listing its mem-



pages show Max Beberman, director of the project, teaching a first-year class how to solve equations. The squares on the blackboard represent spaces for a numeral. They are used instead of *x* early in the course to make clear the meaning of letters in algebra.

bers within braces, *e.g.*, {1, 2, 3, 4, 5}. There is also a conventional shorthand for statements about sets and their members. For example, the statement that 4 is a member of the set A is written $4 \in A$; that 6 is not a member of the set is written $6 \notin A$. To say that A is a subset of B they write $A \subseteq B$. The empty set is "$\varnothing$". The letter $\bar{U}$ denotes a "universal" set—embracing all the members pertinent to a particular discussion; for example, the universal set for plane geometry consists of all the points in a plane.

This unfamiliar language may seem an undue burden to place on beginning students in mathematics, but actually with a little practice it soon becomes easy to read—considerably easier than mastering stenographer's shorthand.

To illustrate the operations on sets I shall mention only three. The union of two sets A and B (written $A \cup B$) is the set consisting of all the members that are in A, in B, or in both. Thus the union of {1, 2, 3, 4} and {2, 3, 4, 5} is {1, 2, 3, 4, 5}. The intersection of two sets ($A \cap B$) is the set of all the elements that are common to the two sets. Thus the intersection of {1, 2, 3, 4} and {2, 3, 4, 5} is {2, 3, 4}. Finally, the complement of a subset A (written $\bar{A}$) is all the members of the universal set that are not in the subset; *e.g.*, if the universal set is {1, 2, 3, 4, 5} and a subset is {2, 3}, its complement is {1, 4, 5}. These ideas can be presented in a graphic way by simple figures known as Venn diagrams [*see page 317*].

Set theory helps to lay bare the unity of mathematics. Algebra and geometry are both concerned with sets—algebra with sets of numbers, geometry with sets of points. The specific operations in both subjects can be considered examples of the general set operations of union, intersection, and so on.

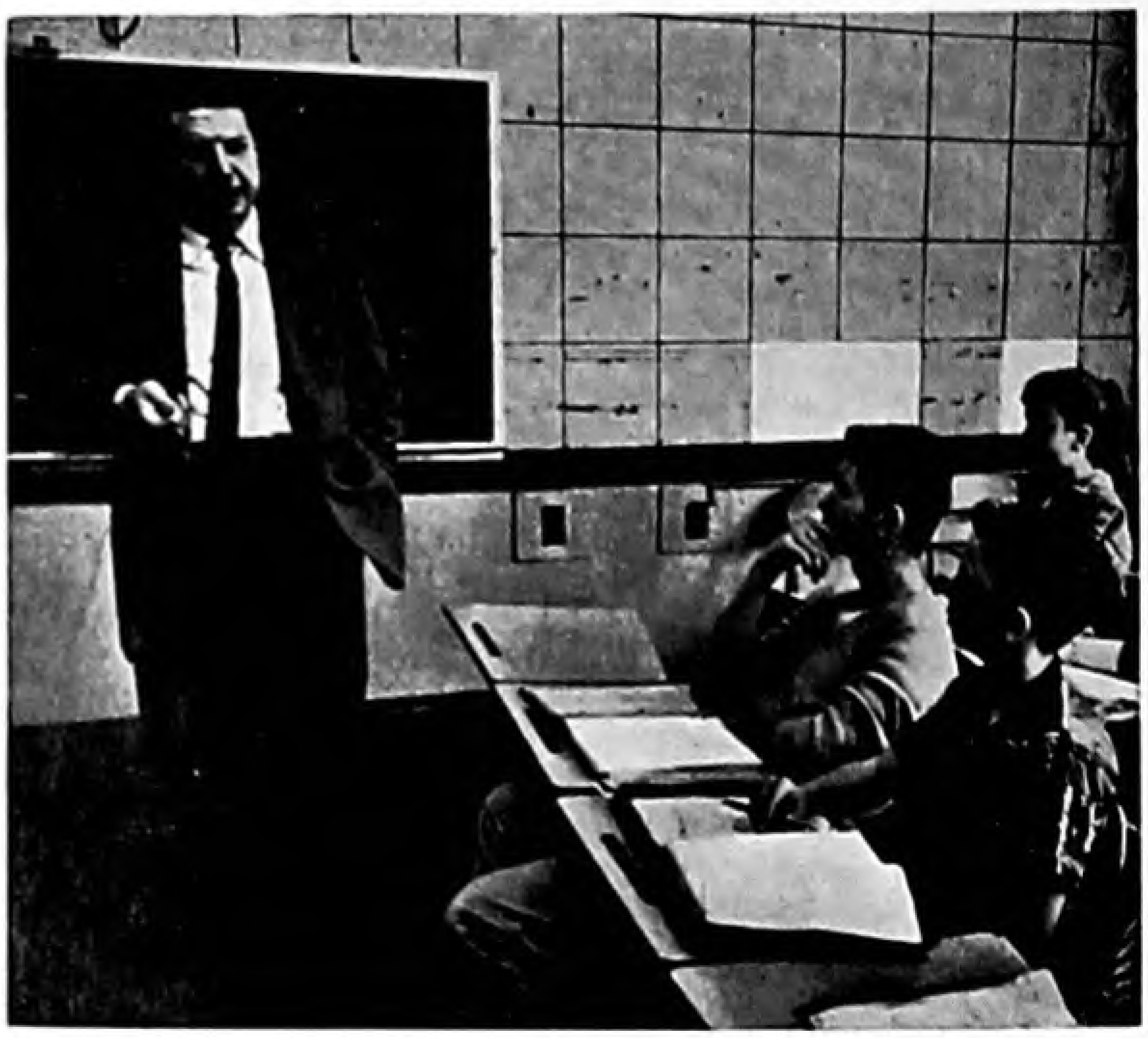
Let us see how the concept of sets can help to clarify the notions of a variable and of an equation in algebra. The College Board group recommends that students be taught first of all that algebra deals with sets of numbers and relations between them. They would then learn that a variable is simply a general name for members of a set. As such it can be represented by a letter. For example, if the set is the series of all the whole numbers, then x can be any whole number, and $x + 1$ can be $1 + 1$ or $2 + 1$ or $3 + 1$, etc. Further, an equation is a statement about a relationship between members of a set. It may be true or false: $3 + 2 = 5$ is true, $3 + 2 = 6$ is false, but both are statements. If a statement contains letters, it is noncommittal: $x + 2 = 5$ is neither true nor false until the place held by x is filled with some member of the appropriate set. Early in the course the students would also learn to work with inequalities, using the symbols $>$ (greater than) and $<$ (less than).

Now any relation can be regarded as a specification for selecting a certain subset from the universal set of numbers under consideration; in set terminology it can be called a set "builder." Thus if the universal set is all the real numbers, the relation $x + 2 = 5$ selects the set {3}, or $x + 2 > 5$ selects the set of all numbers greater than 3. The set selected by a relation is known as its solution set.

The same concept applies to pairs of numbers and can deal with an equation with two unknowns, conventionally represented by x and y. Such an equation of course has more than one solution. For example, the set selected by the equation $5x + 3y = 15$ includes such pairs as (0, 5), (3, 0), (1, 3⅓). Now anyone familiar with coordinate systems recognizes this at once as the wedding of algebra with geometry: a pair of number variables (x, y) can define either a point or a line [*see diagrams on page 318*]. Thinking in terms of sets, a student can see the solution of two simultaneous equations of algebra is a matter of finding the intersection of the two solution sets: the solution is the pair of numbers that is common to both sets.

The notion of sets is particularly helpful in dealing with inequalities. The meaning of an expression such as $5x + 3y > 15$ becomes clearer when it is considered as the selector of those pairs of numbers (or points) that lie above the line $5x + 3y = 15$ on a graph. The expression $x^2 + y^2 < 16$ selects the points inside the circle $x^2 + y^2 = 16$. To "solve" this pair of inequalities simultaneously is again a question of finding the intersection of their solution sets [*see the middle diagram in the second column on page 318*].

All these are merely examples to illustrate the College Board group's approach in its proposed revision of the algebra course. Except for the emphasis on inequalities, it does not appreciably



SEQUENCE IS CONTINUED as Beberman discusses the meaning of equations and of the transformations used to solve them. Some excerpts: Teacher: "How can we be *sure* this equation is true?" Pupil: "If we replace the holes by numbers, er, by numerals, the

change the subject matter of the course, but it offers that subject matter in a new context.

The New Geometry

In geometry the group proposes to change the course radically. In the first place, it wants to eliminate most of the propositions and theorems that students are now required to prove, on the grounds that the students will already have had some training in deductive reasoning in the algebra course and that many Euclidean proofs can be demonstrated more easily by other methods.

The College Board Commission would cut down the theorems to be proved to about 12, in place of the 100-odd in today's geometry books. These 12 would provide a sample of how the facts of geometry can be deduced from a set of axioms. From them the students would develop theorems and facts about triangles, parallel lines, similar triangles and finally the Pythagorean theorem (a triangle is a right triangle if and only if the square of the hypotenuse is equal to the sum of the squares of the two sides). The reason for stopping with the Pythagorean theorem is that it makes possible a shift to analytic geometry—that is, algebraic solution of geometric problems with the help of graphs. The analytic methods show the use of solution sets in geometry [*see diagrams on page 319*].

The Commission does not propose to abandon classical Euclidean proofs entirely. Some problems are in fact easier to solve by the Euclidean procedure than by the analytic method. Students would be encouraged to find and use the most effective approach to each problem.

The recommended short cuts would reduce the time needed to cover plane geometry to less than a year, and the Commission proposes to use the time saved to include some solid geometry. Students would be encouraged to think in three dimensions as well as in two, and they would study the important theorems of solid geometry, mostly from an intuitive point of view.

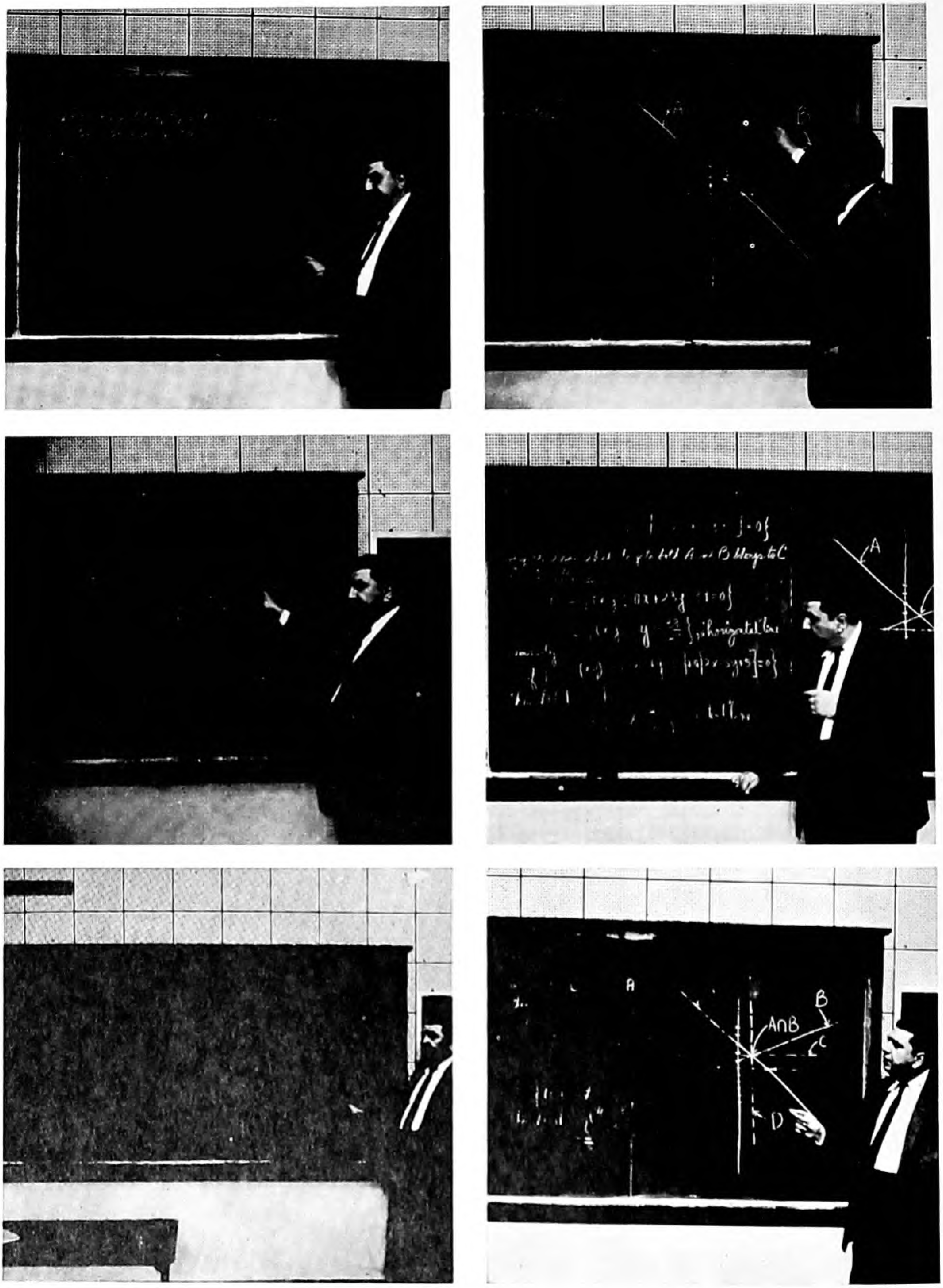
In the third year, after elementary algebra and geometry, the mathematics course would go on to further work in algebra and some trigonometry, with emphasis on the mathematical behavior of the trigonometric functions (sine cosine, etc.) and their application to such problems as the study of vectors rather than the solution of triangles. For the fourth year the Commission would offer students who continued in mathematics two half-year courses. The first, called elementary analysis, would deal with the idea of relations and functions from a more advanced point of view. It would investigate the properties of polynomials (algebraic expressions containing many terms) and of logarithmic, exponential and trigonometric functions. The idea of limits would be introduced informally, and students would learn a little calculus, *i.e.*, how to differentiate and integrate polynomials. In the second half of the year there would be a course in probability and statistical inference. Here a student would learn to deal with sets of scattered measurements or observations by means of statistics. He would learn how mathematics is applied to processes governed by chance. The course would demonstrate some of the methods of computing the reliability of a sampling program and of measuring the statistical significance of results. The Commission points out that this material is probably more closely related to the daily lives of people than any other part of mathematics. Also, statistics and probability are becoming increasingly important in science and industry.

Because most of the topics in the probability course have never been taught in high school, the Commission has prepared a new textbook for this course; it is already being used in a few schools. The book presents the subject first intuitively and then from a more formal mathematical standpoint. Set theory is used extensively in the course. I found the text readable and interesting and the material no more difficult than many of the traditional topics in advanced high-school mathematics.

The Commission also would like high schools to offer, for students who take part in the College Board's Advanced Placement Program, a course in calculus and analytic geometry—the usual college freshman course.

Such, in brief, is the College Board Commission's program for reforming the teaching of mathematics in the nation's high schools. It is urging schools and teachers to try all or any part of its ideas, in the hope that the ideas can be tested to see which will work and which will not. It is confident that once students

two sides of the equation are both names for the same number." Teacher: "What principles did we use in this operation?" Pupil: "We used the commutative principle and added, and we also used the distributive principle to combine the 12 square and 175 square."



IDEAS OF SET THEORY play a major role in the Illinois course. In this sequence Beberman applies them to the solution of simultaneous equations. In the material on the blackboard *A* and *B* represent the sets of number pairs belonging to the two equations. The solution is their intersection. *C* and *D* (*bottom*) are the sets formed by the elimination of *x* and *y* respectively.

have been introduced to the beauty of modern mathematics, the subject will acquire a new vitality in the schools.

Although the Commission's program is merely a proposal, obviously the College Board is in a position to exercise a powerful effect on the high schools, through its examinations and its close relationship with leading colleges. The Commission has already begun to reach teachers throughout the country by means of pamphlets describing phases of its work, and it will publish its comprehensive report this fall. Very probably its suggestions will generate the writing of radically new textbooks within the next couple of years.

The Illinois Program

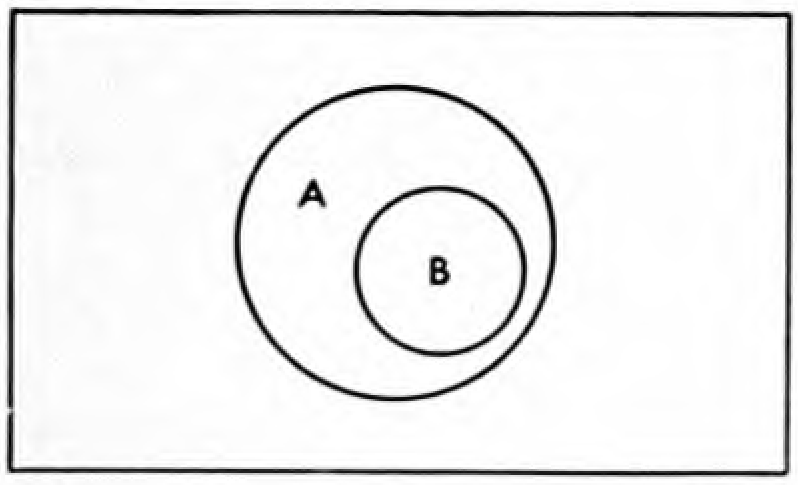
Now let us turn to another approach—the program stemming from the University of Illinois, which has had some public attention in newspapers. The leading spirits of this movement are Max Beberman, a teacher in the University High School at Illinois, and Herbert E. Vaughan, a mathematician on the University faculty. They have worked out a program which goes so far in the direction of modernism that it makes the College Board Commission's program look almost antique.

The Illinois experiment began at the University High School, an adjunct of the University's College of Education, in 1952. It has grown into a four-year program which is being tried this year in a dozen high schools in Illinois, Missouri and Massachusetts, and is also being taught to interested employees of the Polaroid Corporation in Cambridge, Mass. With financing from the Carnegie Corporation, the Illinois group has produced a complete series of textbooks and teaching manuals and brings teachers to the University for up to a year of observation and indoctrination in its courses.

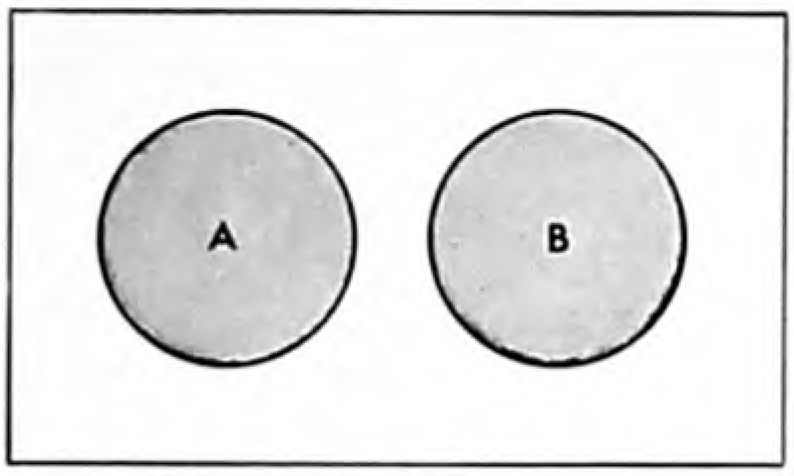
The approach to mathematics via abstract generalizations is the very cornerstone of the Illinois program—not merely an exercise in reasoning or a sampling of some of the foundations of mathematics. Pupils in the ninth grade begin their study of algebra with a set of axioms from which they proceed to prove all the rules they must master. The axioms, called "principles of arithmetic," include such things as the definition of zero (any number times zero equals zero), the definition of the number 1 (any number times 1 equals itself) and of course the commutative, associative and distributive laws. The children are led into their rigorous program by easy stages, sometimes in story form. For example, they learn that a number must be distinguished from the symbol used to represent it by means of a correspondence between Ed Brown, a student at "Zabranchburg High," and Paul Moore, a pen pal in Alaska. The two chums have fallen into the unlikely habit of writing each other about arithmetic. Young Paul advances the reasonable proposition that if you take 2 away from 21, you should be left with 1. Ed is startled: is this a joke, or does Paul have something? The text straightens things out by explaining that "21" is a name for the number, not to be taken too literally. It goes on to show that the number 21 can be described by other names: *e.g.*, "20 + 1" or "7 × 3." The pupils proceed to learn that a letter can stand for a number and that it can be given various meanings (*i.e.*, it is a "variable" or "placeholder" for a number). Indeed, the teacher uses blank boxes for the unknown numbers in equations [*see photographs on pages 312 and 313*]. A large portion of the first-year course is devoted to driving this point home. The text observes that a letter plays the same role in a mathematical statement that a pronoun does in ordinary language. The statement "x + 2 = 6" is like the sentence "He was a president of the U. S." Neither is true nor false until a name is put into the proper place. The teacher calls the letters of algebra "pronumerals," and uses this term throughout the course.

A good part of the first year is also given to introducing and exploring the idea of sets. Because of the long pronumeral introduction and the careful development of the set concept, the course covers less ground than the traditional course in elementary algebra.

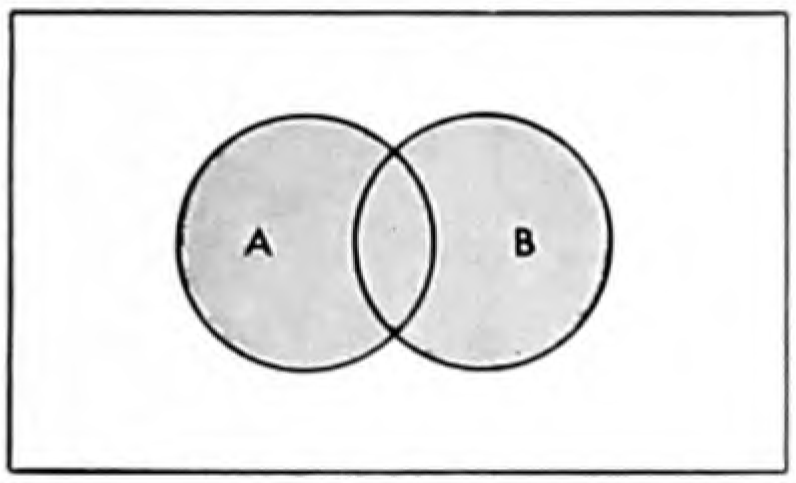
In the second year the program goes modern for fair. This course is nothing less than "a development of Euclidean geometry which is as rigorous as, for example, that due to Hilbert, and yet which is, we believe, accessible to students who have mastered the first course." It attempts to make clear the nature of geometry as a pure deductive theory which, in itself, has only logical structure and is empty of content until specific interpretations are introduced. Consider, for instance, three postulates which are necessary for deducing the rules of Euclidean geometry. These are: (1) every line is a set of points and contains at least two points; (2) there are three points which do not belong to the same line; (3) every two points any-



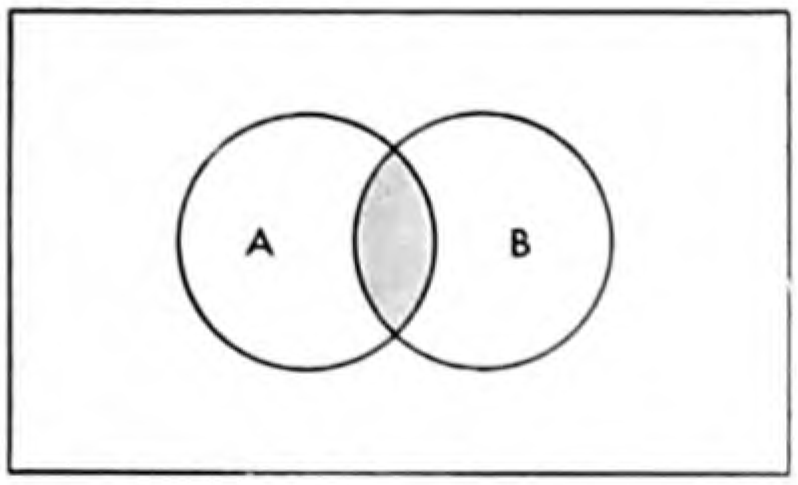
B ⊆ A



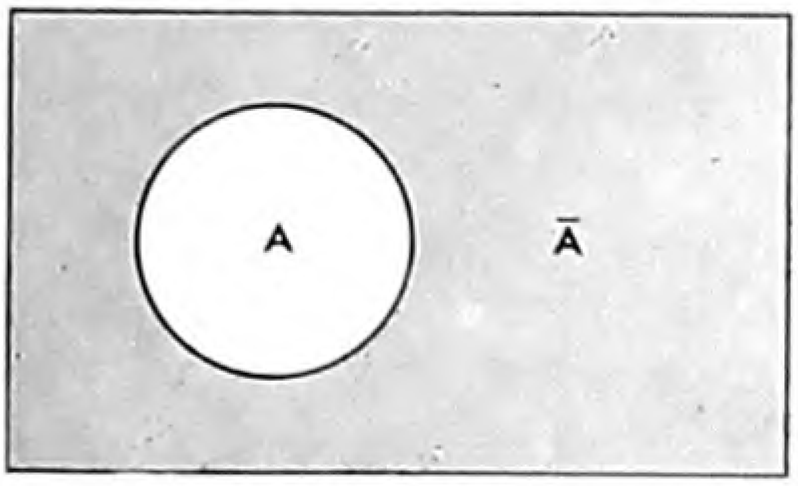
A ∪ B



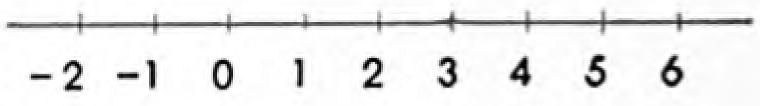
A ∪ B



A ∩ B

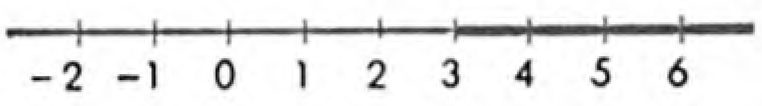


VENN DIAGRAMS are simple pictorial representations of set relations. Top diagram expresses the idea that *B* is a subset of *A*. Shading in second and third diagrams indicates the union of *A* and *B*. In fourth diagram it indicates their intersection. At bottom it represents the complement of *A*.



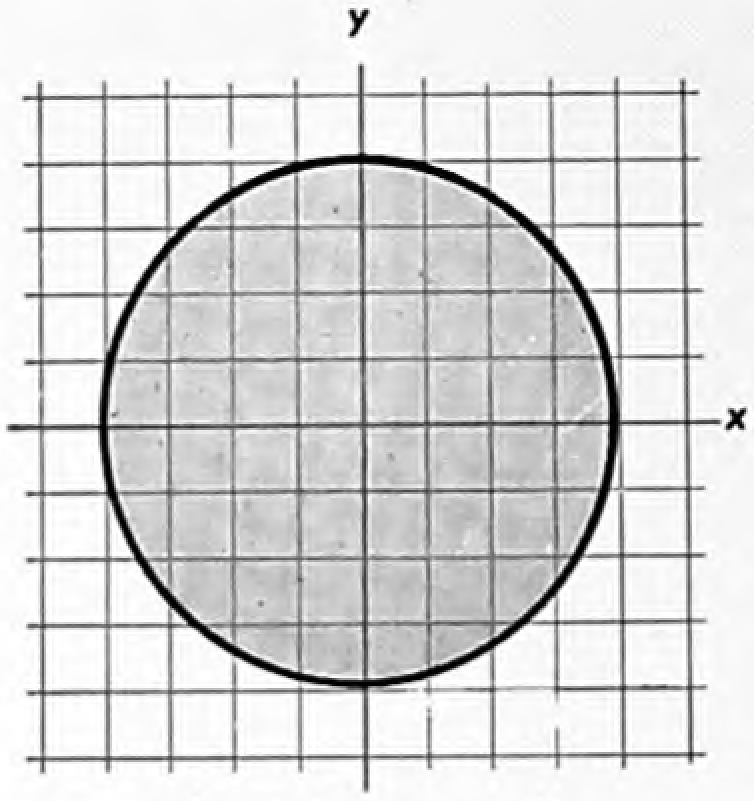
$\{x \mid x + 2 = 5\}$

SETS ARE REPRESENTED geometrically as well as algebraically. Dot represents the set described by equation $x + 2 = 5$.



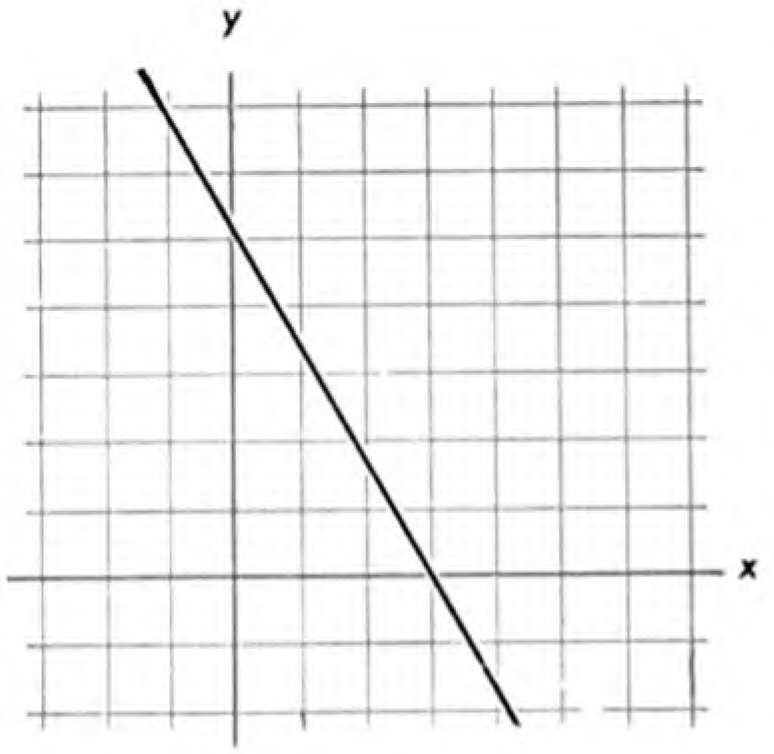
$\{x \mid x + 2 > 5\}$

INEQUALITIES also describe sets. The heavy section of the line in this diagram is the infinite set described by $x + 2 > 5$.



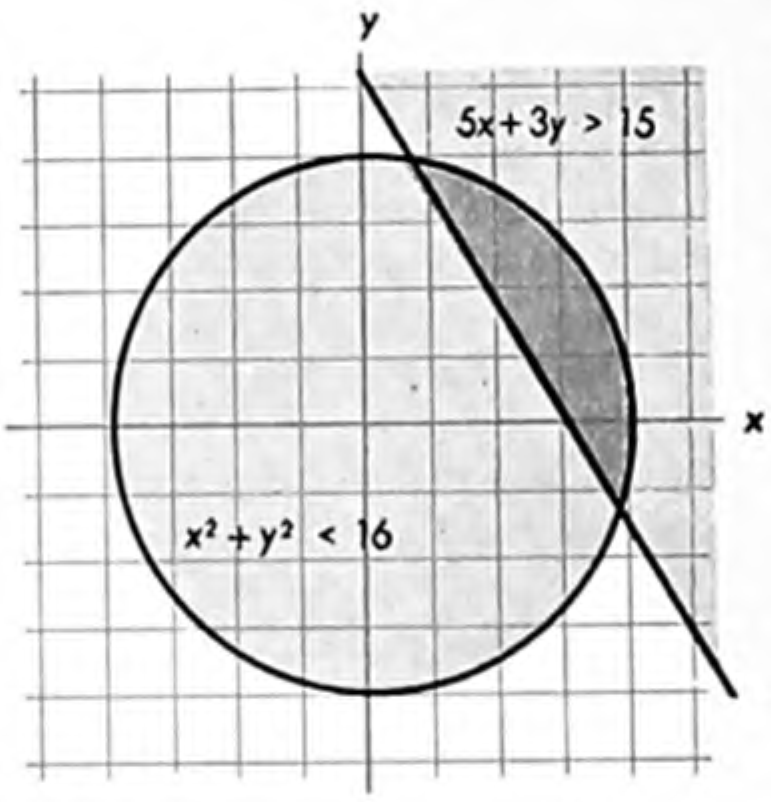
$\{x, y \mid x^2 + y^2 < 16\}$

NUMBER PAIRS that belong to the set selected by the inequality appearing above are graphed as the points inside a circle.



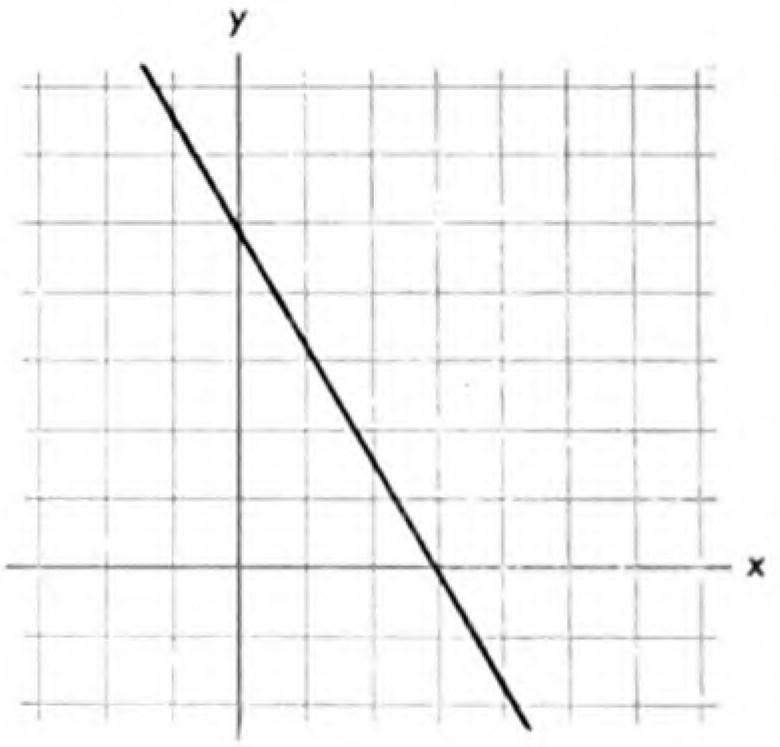
$\{x, y \mid 5x + 3y = 15\}$

TWO-VARIABLE expressions describe sets of pairs of numbers. Points on this line represent pairs belonging to $5x + 3y = 15$.



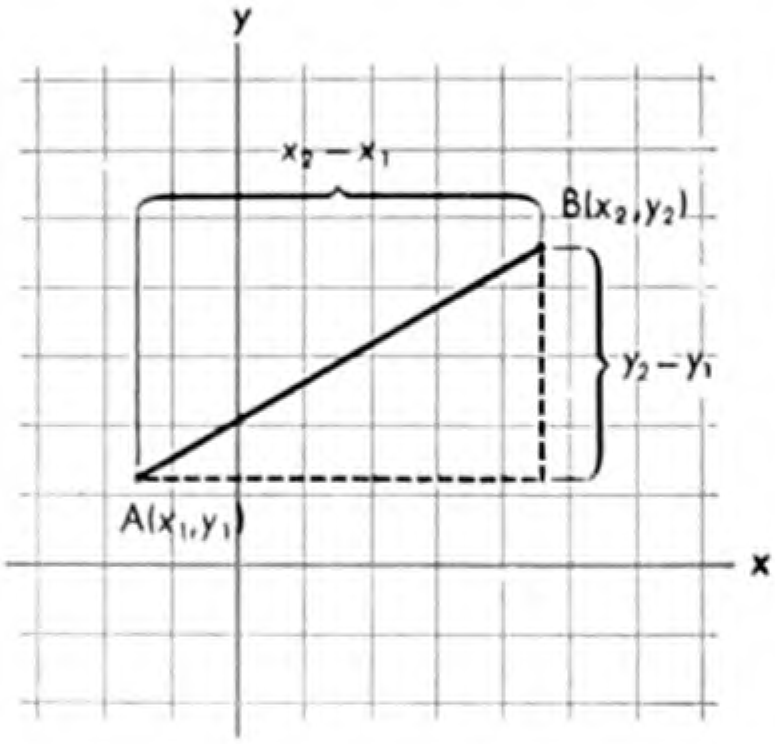
$\{x,y \mid x^2 + y^2 < 16\} \cap \{x,y \mid 5x + 3y > 15\}$

INTERSECTION of sets described by two inequalities is represented graphically by the heavily shaded area in this diagram.



$\{x, y \mid 5x + 3y > 15\}$

SET-BUILDER idea helps give meaning to inequalities. Shaded area represents set selected by the inequality in brackets.



$AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

DISTANCE FORMULA in analytic geometry appears under this diagram. Lower-case letters are coordinates of end-points of line.

where are contained in a line. The teacher brings out that the words "point" and "line" can be interpreted in various ways. The main model used in the course is the "number plane," where points and lines are defined in terms of pairs of numbers (x, y); the course develops all the necessary postulates of plane geometry from this model. But to emphasize that the deductive system can have various concrete interpretations, the teacher examines several models, including one in which "point" stands for businessman and "line" for a partnership. The postulates apply here just as they do in geometry. Suppose there are three businessmen, *A*, *B* and *C*, each of whom forms a two-man partnership with each of the others. Each partnership (*AB*, *AC* and *BC*) is a set of businessmen and contains at least two members (postulate 1). None of the partnerships contains all three businessmen (postulate 2). Every two businessmen are contained in a partnership (postulate 3).

I can make a little clearer how the teacher and children deal with these matters by quoting a recording of a review discussion of postulate systems between Beberman and a bright class:

Teacher: "Where do these postulates come from?"

George: "From the number plane."

Teacher: "What do you mean by that?"

George: "They are properties of the number plane."

Teacher: "If we're talking about the number plane when we say, for example, that each line is a set of points and contains at least two points, is that statement true or false?"

George: "True."

Teacher: "But suppose we are *not* talking about the number plane. Then what about these postulates?"

George: "False."

Chorus: "No!!"

Jim: "You have to give a specific meaning to 'line' and 'point' before you can tell whether they are true or false."

Teacher: "What is a 'model' of a postulate system?"

Jane: "Something that has the properties expressed by the postulates or that satisfies the postulates."

Teacher: "Suppose there are several interpretations for the postulates. You can be talking about the number plane, or about businessmen and corporations, or about class presidents and committees. So what?"

George: "Well, when you try to deduce a theorem from the postulates, you can use a model to find out in some cases

that you can't deduce it because it's false for that model."

In the hands of a teacher like Beberman the discussion in a better-than-average class is alert and spirited. A leaflet put out by the Illinois group asserts: "High-school students have a profound interest in *ideas*. They enjoy working with abstractions. . . . Despite the current fashion to point out the usefulness of mathematics in various occupations, most high-school students are not genuinely stirred by such a 'sales campaign.' The goal of vocational utility is too remote to make much difference to a ninth grader. He wants to know how mathematics fits into his own world. And, happily, that world is full of fancy and abstractions. Thus students become interested in mathematics because it gives them quick access to a kind of adventure, which is enticing and satisfying."

The course is frankly experimental. Some of the material has proved too time-consuming or too hard to teach. At the moment the Illinois group is considering whether it should postpone some of the rigorous development to the fourth year. But Beberman starts with the attitude that you don't know what you can teach children until you try.

It is too early to tell how much the pupils eventually get from the course; the first group is just completing the four-year program this term. But I have visited the University High School classroom and can testify that the students seem to carry their burden cheerfully and even enthusiastically. The big question is: How would such a course go generally? Some critics concede that it may work well with a gifted teacher and bright students but strongly doubt that the average teacher or the average class could handle it successfully.

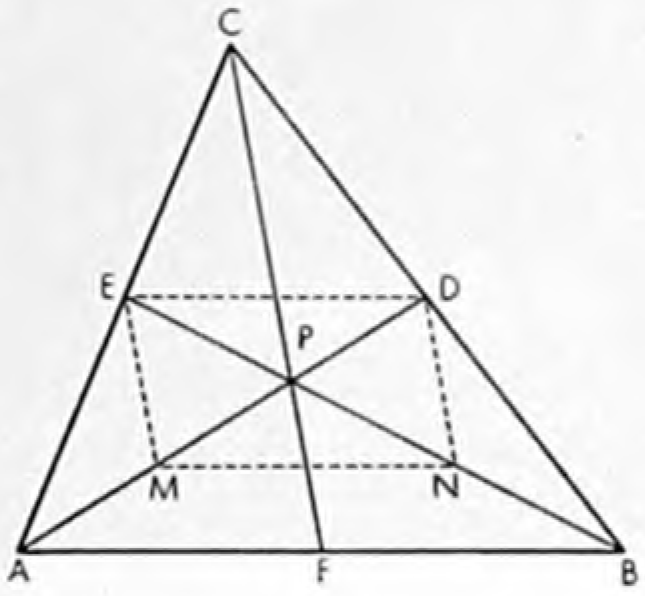
The Critics

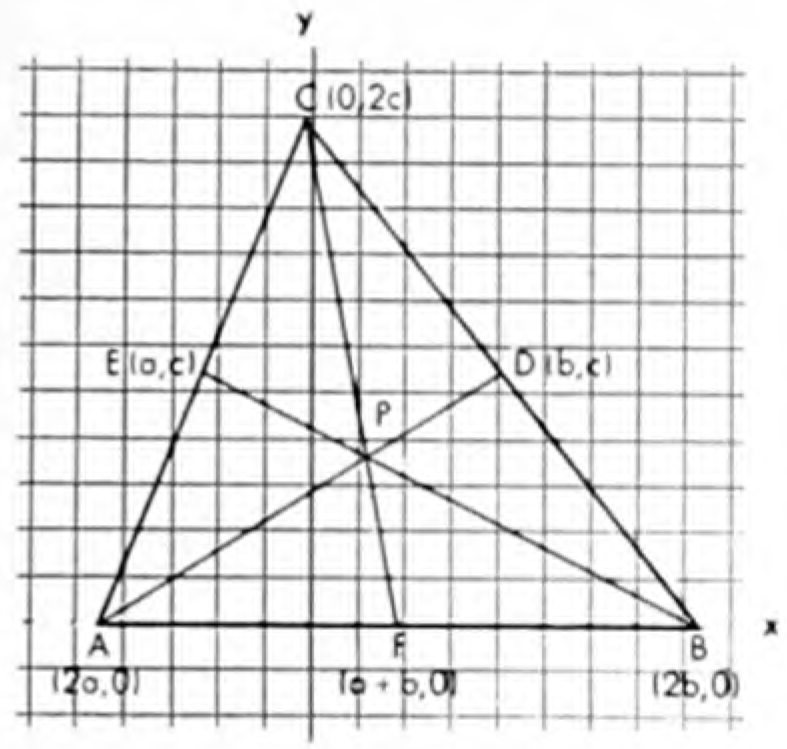
There are those who believe that the whole approach of the modernists is fundamentally wrong. One of the most articulate of these critics is Morris Kline, a professor at New York University's Institute of Mathematical Sciences and a popular writer on modern developments in mathematics. He doubts that the abstract approach will get youngsters interested in mathematics.

Kline argues that you do not abstract until you know what you are abstracting from. Professional mathematicians did not arrive at an understanding of the abstract general features of mathematics until after they had explored many of its specific branches. Why should a schoolboy be expected to? Moreover, Kline believes that set theory has no place in an elementary curriculum: its applications there can only be trivial, and even its importance in advanced mathematics is overemphasized.

Kline and others hold that the way to stimulate more interest in mathematics is to make the teaching more concrete rather than more abstract—to relate it more vividly to problems of the real world. Kline would introduce mathematical ideas by means of simple physical experiments and examples drawn from fields such as music and other arts. An active advocate of this sort of approach is the English mathematician and teacher W. W. Sawyer, now at the University of Illinois. Sawyer has built a number of simple devices which he uses to illustrate mathematical ideas [*see photographs on next page*].

But the anti-modernists seem to be in a losing cause. They are not organized, nor have they formulated a specific program. Meanwhile the modernist movement is spreading rapidly. At the college level, a number of colleges have adopted a new freshman course designed by a committee of the Mathematical Association of America. The course consists of a half year of calculus and a half year on modern topics, collectively called "discrete" mathematics. It deals with noncontinuous sets (*i.e.*, of discrete numbers or objects). This branch of set theory is finding increasing application in science, particularly social





Let AD and BE intersect at P.

Mark M and N, mid-points of AP and BP respectively.

ED is parallel to AB and equal to ½ of it.

MN is parallel to AB and equal to ½ of it.

∴ ED is parallel and equal to MN.

∴ MP = DP and EP = NP.

But AM = MP and BN = NP.

∴ P is ⅔ of the way from A to D and from B to E.

By the same construction, using AD and CF, their intersection can also be shown to be ⅔ of the way from A to D.

∴ the medians intersect in a common point which divides them in the ratio of 2:1.

Q.E.D.

By the rules for writing equations of lines, the equations of AD, BE and CF are respectively

$$cx + (2a - b)y = 2ac$$
$$cx + (2b - a)y = 2bc$$
$$2cx + (a + b)y = 2ac + 2bc$$

These three equations have the common solution

$$x = \tfrac{2}{3}(a + b),\ y = \tfrac{2}{3}c$$

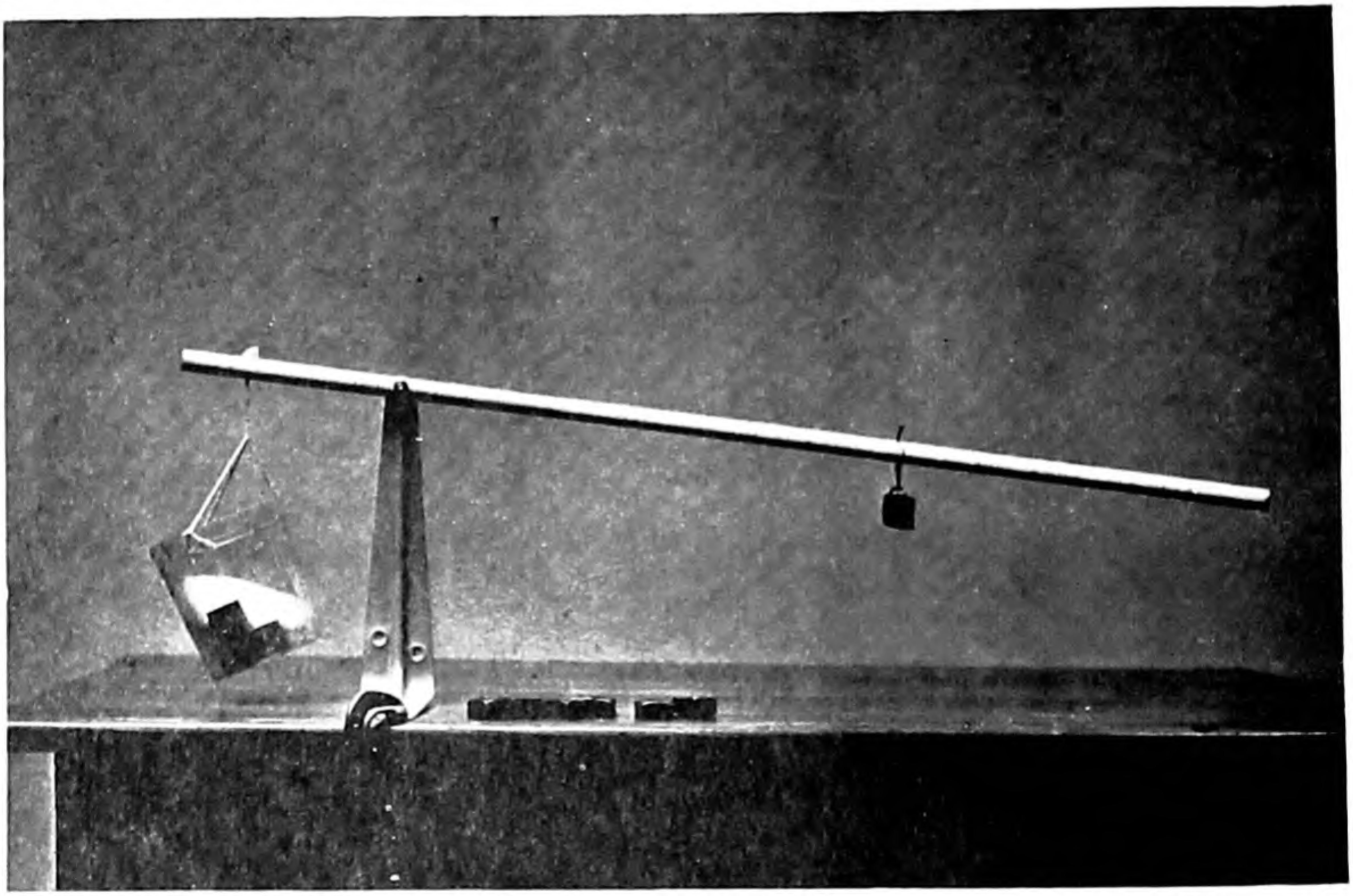
∴ the medians meet in a common point.

Substituting the proper coordinates in the distance formula shows that

AP = 2PD, BP = 2PE and CP = 2PF

Q.E.D.

PROOF OF THEOREM (the medians of a triangle meet in a common point which divides them in the ratio 2:1) is done by Euclidean method (*left*) and analytic geometry (*right*).



SIMPLE MECHANICAL DEVICES are used by mathematician W. W. Sawyer to illustrate mathematical principles. At top is a water clock whose design involves calculations of the relation between the rate of flow of liquid and its level in the vessel. At bottom is a balance for which students discover a relation between position of the counterweight and number of nuts in weighing pan.

science, which deals typically with groups of persons, units of product, etc. Even physics is discrete at the atomic level, and matrix algebra, one of the topics in discrete mathematics, lies at the foundation of quantum theory. The new college course is intended to give liberal arts students a taste of some modern mathematical ideas and also to acquaint social science majors with mathematical techniques. The text now in most common use includes symbolic logic, probability, the algebra of vectors and matrices, the theory of games and linear programming.

The Teachers

Mathematics teaching in the nation's high schools and elementary schools obviously will not be changed overnight. Most teachers are not prepared to teach either the modern material or any other new scheme. However, the current great expansion of teachers' summer institutes and in-service courses, financed by the National Science Foundation, private corporations and others, is a golden opportunity. This summer there will be 10 university institutes in mathematics, all teaching mainly the modern topics. Some 40 other universities are offering courses in both mathematics and science, and most of them are stressing the modern approach. It appears likely that most U. S. high schools will have shifted more or less toward the modern approach within five years or so.

Finally I must mention a new project which in the long run may have a stronger impact on mathematics teaching in the U. S. than any of the programs discussed in this article. It is the School Mathematics Study Committee, set up by the mathematics department of Yale University and patterned after the Physical Science Study Committee at the Massachusetts Institute of Technology described in this magazine [see "The Teaching of Elementary Physics," by Walter C. Michels, Offprint 229]. The new group will enlist a large number of mathematicians and teachers to consider mathematics teaching in high schools and junior high schools, and very likely it will prepare an elaborate program of textbooks and other teaching aids, as the physical science group is doing. This summer the Committee will hold a conference to review the present experiments and the whole problem. What it will come up with is anyone's guess, but in manpower and financial support it may well carry more weight than any other group seeking to revitalize the teaching of mathematics.

The Author

E. P. ROSENBAUM is a member of the board of editors of SCIENTIFIC AMERICAN. He approaches the subject of his article not only as a journalist but also as a teacher: for several years he taught mathematics and physics at the Milford School, a preparatory school in Connecticut.

SPECIAL NOTE TO TEACHERS: Each article in this volume, plus more than 660 others, is available as a separate, self-bound SCIENTIFIC AMERICAN Offprint. Offprints may be ordered in any combination and in any quantity. Teachers who want to adopt articles for their courses, therefore, can ensure that each student has his own set. Students' sets are collated by the publisher before shipment.

SCIENTIFIC AMERICAN September 1953, Vol. 189, No. 3, pp. 64-70 OFFPRINT 239

WHERE DO COSMIC RAYS COME FROM?

by Bruno Rossi

These high-energy particles may come from the sun, stars or galactic nebulae. A solution to the question would fill in the composition of the universe.

The earth is under a ceaseless rain of particles from space. These cosmic rays, our only material contact with the vast universe outside our planetary system, have excited wonder and eager study ever since they were first discovered 41 years ago. They fall upon us with energies far beyond anything that can be produced on earth. They shatter the atoms of matter and make their nuclei explode into strange fragments. It is the investigation of cosmic rays that has been responsible for the discovery of so many new elementary particles in the past quarter-century: the positron, the various mesons, the V-particles and others which are being discovered even as these lines are written. Besides this, cosmic rays are of great interest in biology, for by producing mutations in genes they are said to have played, and continue to play, a large role in the evolution of life on the earth.

Thus the cosmic rays have been very useful to science. But the big question remains: Where do they come from, and how do they get their fantastic energy?

At six o'clock on the morning of August 7, 1912, a balloon took off from a field near the Austrian town of Aussig. It carried three men, one of them a young physicist named Victor Hess, and three sensitive ionization meters. Hess was out to learn something about the source of a certain mysterious radiation which physicists had been detecting for some time with laboratory instruments. His balloon rose to 16,000 feet, and he found the radiation much stronger there than at sea level. After analyzing his readings, he announced: "The results of my observations are best explained by the assumption that a radiation of very great penetrating power enters our atmosphere from above. . . ."

This was the first recognition of what the U. S. physicist Robert Millikan later named cosmic radiation. The fascinated investigation that ensued concerned itself first of all with finding out what the cosmic rays were.

Outside the earth's atmosphere cosmic radiation consists mainly of protons (nuclei of hydrogen), varying widely in energy. There are few, if any, protons of energy below one billion electron volts (Bev). Most of them are in the range of one to 100,000 Bev. Occasionally a cosmic-ray particle hits the atmosphere with much higher energy, up to 100 million Bev; it produces a gigantic shower containing millions of particles. For comparison, recall that the most powerful accelerator made by man, the Brookhaven Cosmotron, accelerates protons to an energy of a little more than two Bev.

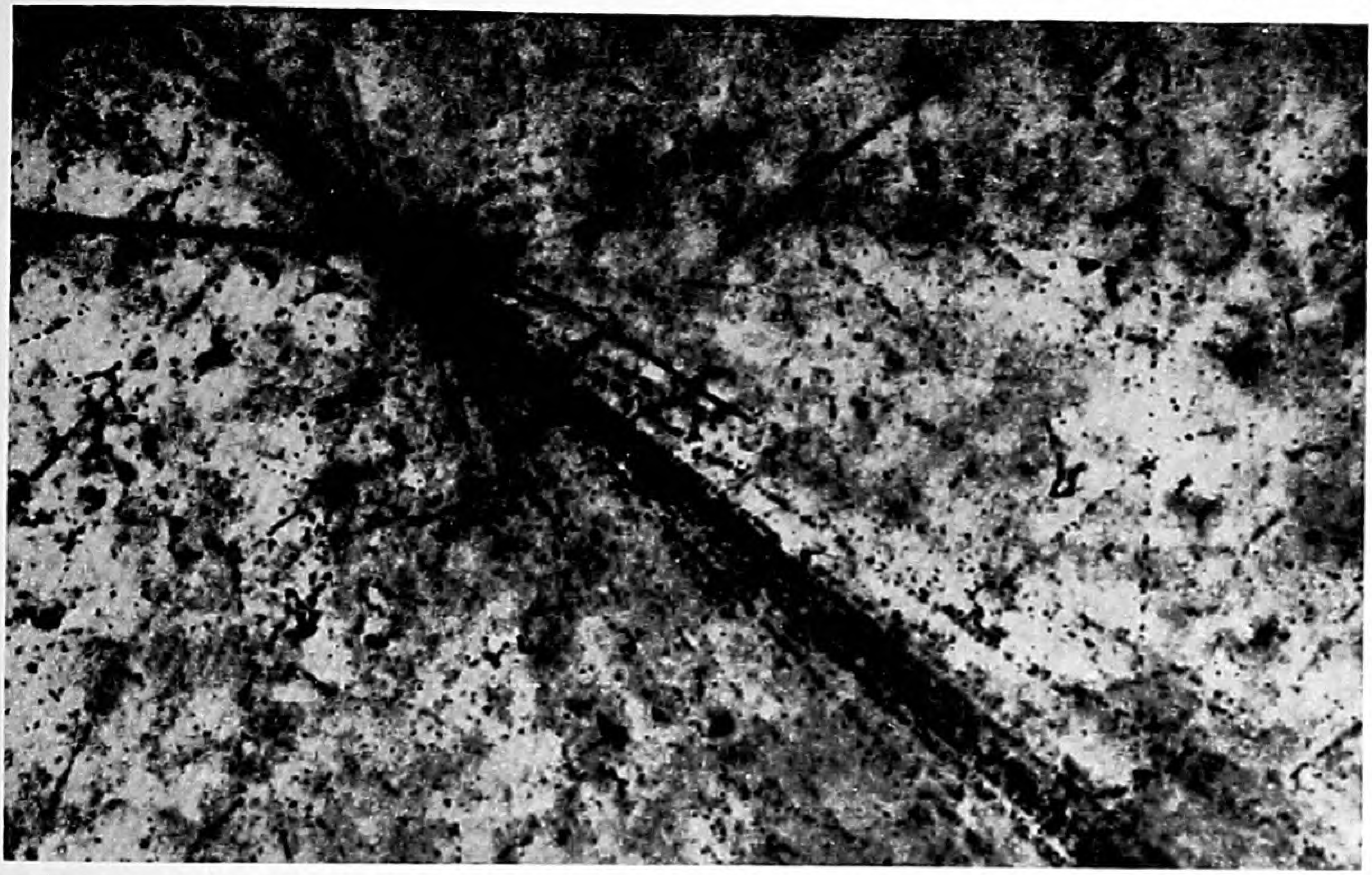
Cosmic rays also contain nuclei of helium and of heavier elements. According to our still incomplete information, the velocity range of all the types of nuclei is approximately the same. At any given velocity, we find approximately 85 helium nuclei and six heavier nuclei for every 1,000 protons. It is interesting to note that the relative abundance of the various nuclei in cosmic rays corresponds closely to the relative abundance of those elements in the universe. Long before striking our atmosphere and beginning the series of collisions by which their energy is eventually dissipated, cosmic-ray particles are deflected by the earth's magnetic field. Some of them are thrown back into space; others reach the earth from a direction which may differ considerably from their original path.

Suggestions as to where the cosmic rays may come from divide themselves into two general schools of thought. One school holds that the cosmic rays were created several billion years ago in a tremendous explosion that gave birth to the universe; since then they have been traveling through space along trajectories curved by the universe's general gravitational field. The trouble with this point of view is that it confines the entire problem to the realm of pure speculation. It therefore seems more profitable to explore the second hypothesis, which is at least theoretically verifiable. This point of view assumes that cosmic rays are produced continuously somewhere in the system of stars which forms our galaxy.

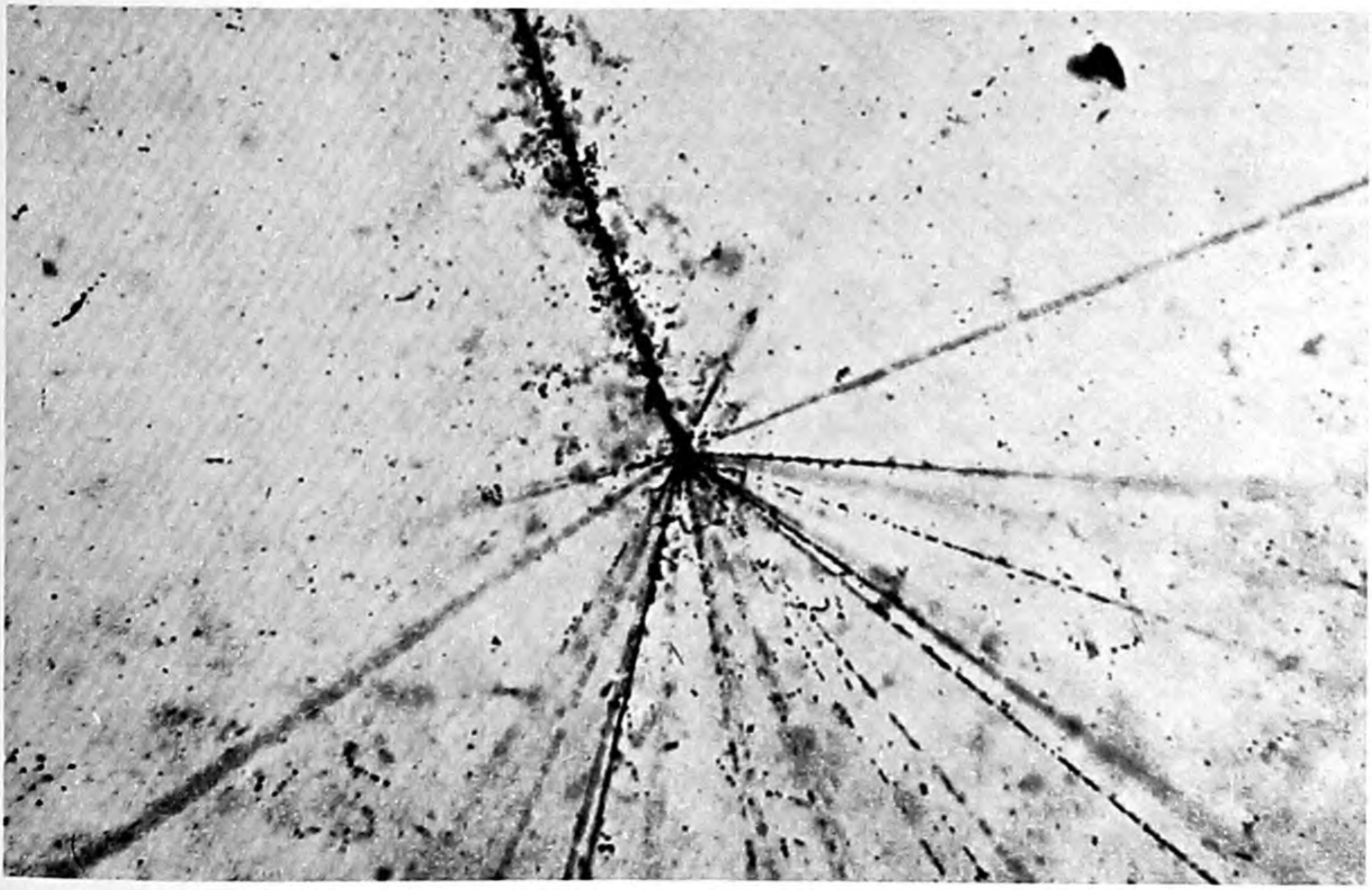
The most attractive theory is that they come from the nearest star, our sun, for the farther away we place the source, the harder it is to account for the relatively heavy intensity of the cosmic-ray fall on the earth. The energy of this fall is very small compared to the energy of the light and heat we get from the sun, but it is comparable to the light from the distant stars.

Correlations between the intensity of cosmic radiation and activity in the sun have, in fact, been observed. Shortly after the appearance of a large flare on the sun, there is sometimes a sudden burst of extra cosmic radiation at the earth. Three such events are on record. On November 19, 1949, cosmic-ray meters at widely separated stations registered abnormally high intensities about one hour after a solar flare had reached its maximum. A detector of cosmic-ray neutrons at Manchester, England, went off the scale. At Climax, Col., a detector of cosmic-ray mesons registered a 180 per cent increase.

Solar flares are often followed by perturbations of the magnetic field around the earth (magnetic storms), which



COSMIC PARTICLE caused this tiny explosion revealed by a photomicrograph of a photographic emulsion especially prepared to record nuclear events. An exceptionally energetic helium nucleus struck a heavier nucleus in the emulsion, giving rise to the jet of mesons going off to the lower right. These photographs were provided by Herman Yagoda of the National Institutes of Health.



PRIMARY COSMIC PARTICLE entered this emulsion from the upper left and shattered a nucleus into about 40 nucleons and mesons. Some of the particles went off at large angles to the plane of the emulsion, and hence are not visible. The nuclear charge of the primary was about 20, that of calcium. The event was recorded during the flight of an unmanned balloon from Pyote, Texas.

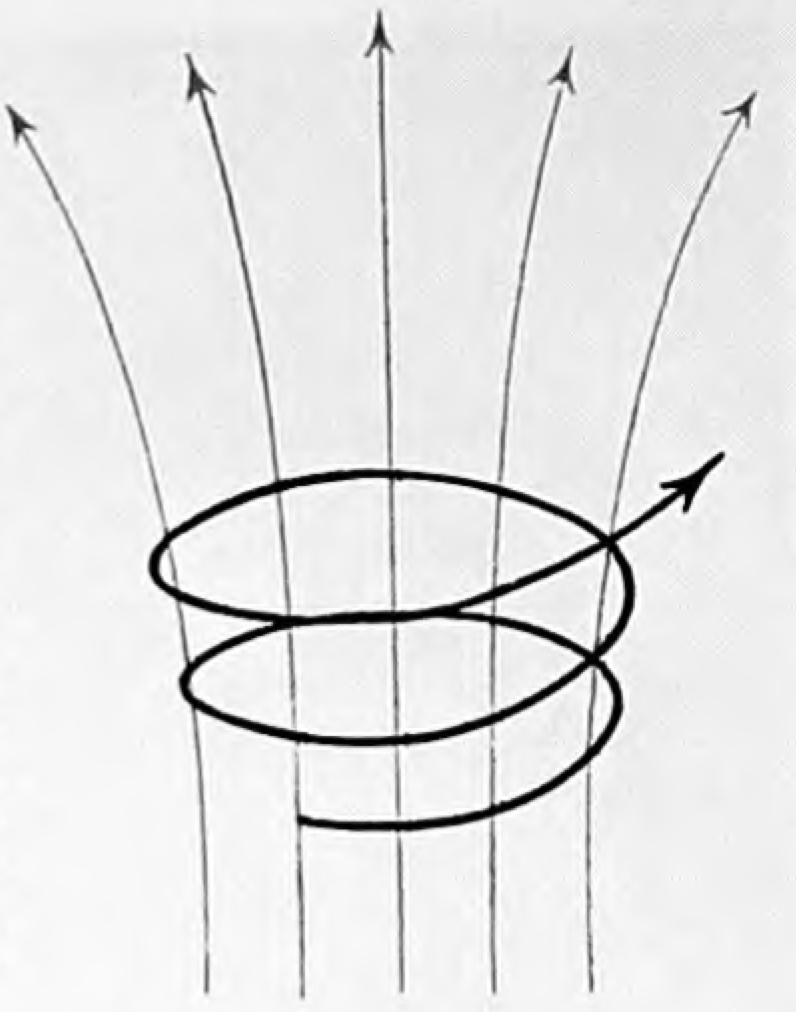
could conceivably modify the flux of cosmic rays reaching the earth even if they came from another source. But the rise just mentioned is much too large to be explained in this manner. It is difficult to escape the conclusion that on November 19, 1949 (and on two similar occasions) the earth was struck by a particularly intense burst of cosmic rays emitted directly from the sun.

This does not necessarily mean that the sun is the sole source of cosmic rays. If they came only from the sun, their intensity should vary markedly with the position of the sun in the sky. Experiments have shown, however, that the variation during the day's 24 hours is less than half of one per cent. Of course, the earth's magnetic field makes cosmic-ray particles of moderate energy describe very complicated trajectories and it thus produces a more or less random distribution in their direction of arrival. But the magnetic field has no appreciable effect on particles of higher energies, and yet they too appear to arrive with almost equal intensity from all parts of the sky.

To answer this objection, Edward Teller of the University of California has suggested that in the region of space surrounding the sun there may exist irregular magnetic fields of sufficient strength to prevent or greatly retard the escape of cosmic rays. This trapping field would scatter cosmic-ray particles back and forth for some time, destroying any original preferred direction of motion. Only on rare occasions would a cosmic-ray beam arrive at the earth directly, as, for example, following the appearance of solar flares.

The Swedish astrophysicist Hannes Alfvén has proposed a possible mechanism to account for the magnetic fields postulated by Teller. It is well known that when a conductor moves in a magnetic field, electric currents are induced; these tend to slow down the motion and at the same time produce a supplementary magnetic field. Thus kinetic energy is exchanged for magnetic energy. Now it is known that the sun continuously ejects clouds of ionized and therefore highly conducting gas. The clouds move outward with tremendous velocities, sometimes approaching one per cent of the velocity of light. In passing through the sun's magnetic field a cloud transforms part of its kinetic energy into magnetic energy, and the magnetic field thus produced is carried along by the cloud in its motion away from the sun.

But even if some such mechanism accounts for the dispersed trajectories of particles coming from the sun, the question of how much that body actually contributes to the observed cosmic-ray flux is still open. Most physicists agree that the sun cannot produce any appreciable fraction of the more energetic particles. In the first place, the radiation observed on the occasion of solar flares contains only particles of moderate energy. This is proved by the fact that the intensity does not rise at the equator, where the energy necessary to penetrate the earth's magnetic field is highest (about 15 Bev for protons). Moreover, the trapping field could not be sufficiently strong to keep particles of the highest observed energies confined to the vicinity of the solar system. It is perfectly possible, however, that cosmic-ray particles of energies below 10 Bev come in part from the sun, and the next question to be considered is the way in which they may be produced.



BETATRON EFFECT of the increasing magnetic field of a sunspot (colored lines) causes particles to accelerate in a spiral.

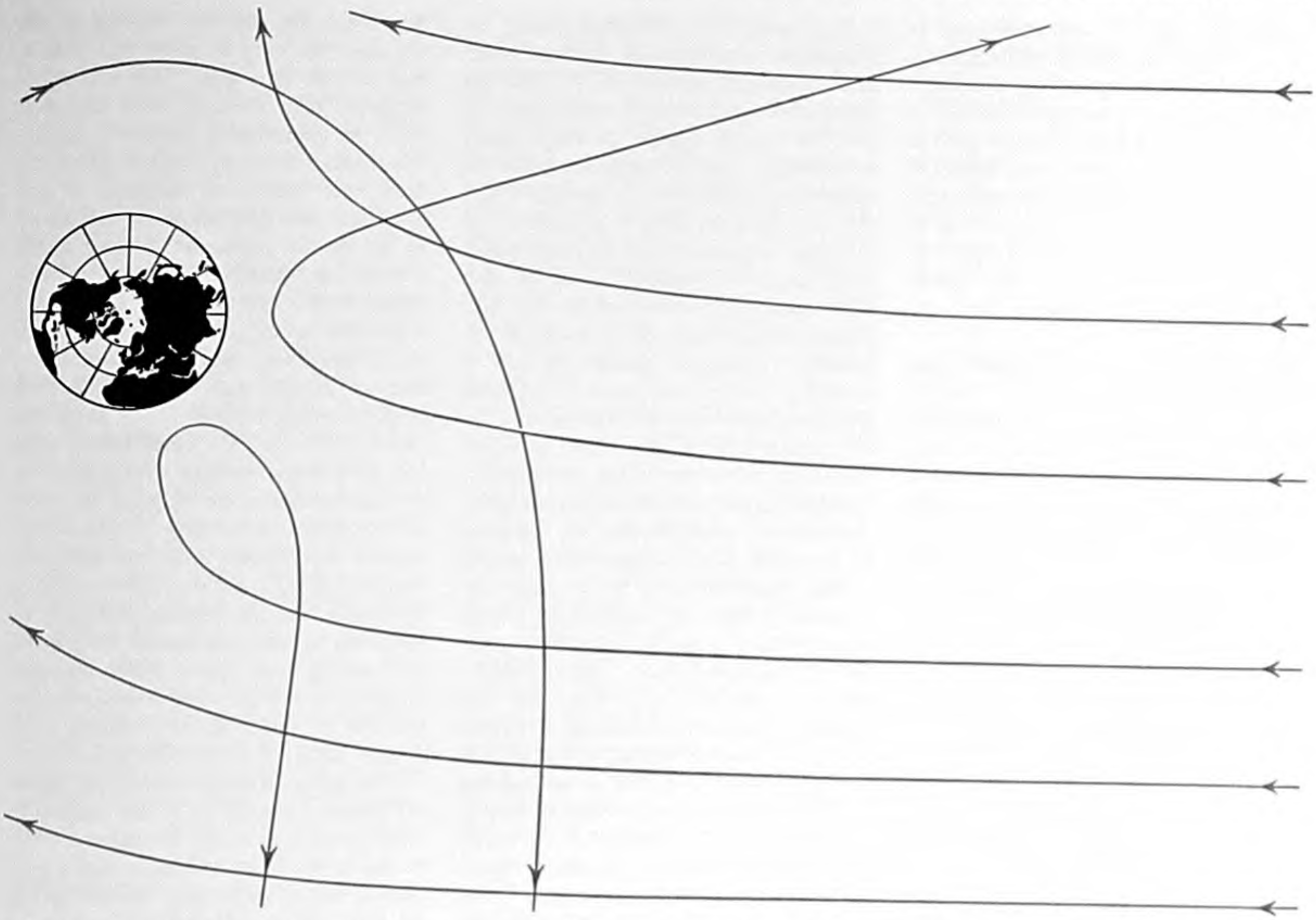
Large-scale electromagnetic phenomena are a part of the sun's activity. One can imagine various situations in which such phenomena could accelerate any charged particles present in the solar atmosphere. A specific process, still regarded as a possible source of cosmic rays, was suggested by W. F. G. Swann 20 years ago.

The sun's surface is often marked by the disturbances called sunspots, which cover areas many thousands of miles in diameter. A sunspot, seen as a dark region on the sun's disk, is the seat of a magnetic field which increases gradually in strength as the spot develops and then decays as the spot fades. A changing magnetic field produces an electric field, according to the well-known laws of electromagnetic induction. That is the basic principle of operation of the betatron, a machine capable of accelerating electrons to energies of several hundred million electron volts. One may crudely describe a betatron as a transformer in which the secondary winding has been replaced by an evacuated tube shaped like a doughnut and containing a source of electrons. As the magnetic flux increases, the electrons are set in motion by the induced electric field, while the magnetic field itself obliges the electrons to follow a circular path inside the doughnut.

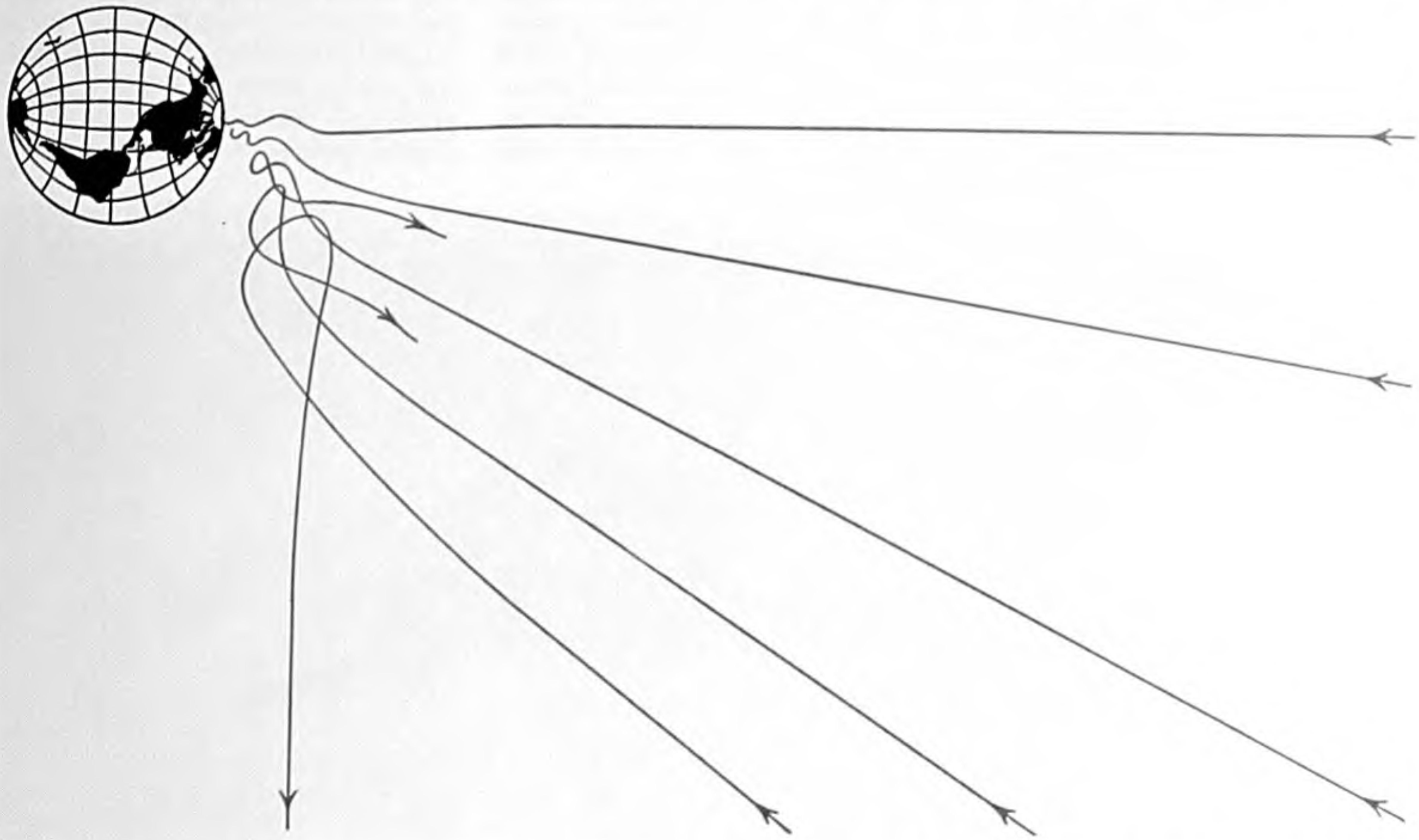
According to Swann's hypothesis, a sunspot operates like a gigantic natural betatron. (Note, however, that the betatron had not yet been developed when this hypothesis was first advanced.) Ions of hydrogen and of other elements in the sun's atmosphere are caught, so to speak, in the electromagnetic vortex of the sunspot. They speed along spiraling paths with increasing velocities and are finally projected into space like stones from a slingshot. The magnetic fields of sunspots could accelerate protons and heavier nuclei up to energies of the order of 10 Bev.

Donald Menzel and W. W. Salisbury, when they were working together at Harvard University, proposed a different mechanism for the acceleration of cosmic-ray particles in the neighborhood of the sun. They based their speculation on the assumption that the sun emits radio waves of very large wavelength. Presumably these waves arise from mechanical turbulence in the highly ionized gases that form the solar atmosphere. Any charged particle which finds itself along the path of a wave is accelerated by the electric field of the wave in a direction at right angles to the path of the wave. However, as soon as the particle acquires some velocity, the magnetic field associated with the wave deflects the particle in a forward direction. By this means a particle may build up an energy of 10 or 100 Bev. Since the particle always lags behind the wave, it will find itself after a while in a position where the electric field of the wave is opposite to its velocity and it will begin to lose the energy it has accumulated. But if the wave field is disturbed at the right moment, as could easily happen under solar conditions, the particle might escape with its maximum energy.

These two theories have a common difficulty. The space around the sun



POLAR PROJECTION of the earth shows some typical trajectories (*red lines*) of charged particles such as the cosmic rays in the plane of the equator. The earth's magnetic field causes some of the particles to veer sharply, others to describe complex loops.



EQUATORIAL PROJECTION shows some typical trajectories of charged particles in the plane of the poles. Because the lines of force in the earth's magnetic field bend downward at the poles, more particles reach the earth's surface there than at the equator.

contains strongly ionized gases and is therefore a good conductor of electricity. One cannot produce a strong electric field in a conducting medium, just as one cannot produce a high voltage with a transformer whose secondary winding is short-circuited. What the theories must provide, therefore, is some mechanism by which a suitable region of space is swept clean of the ionized gas before the particles are accelerated.

Meanwhile, since it is unlikely that the sun contributes more than a modest fraction of the total cosmic radiation observed at the earth, we must continue to look for other possible sources of cosmic rays. If we are to keep the required strength of the cosmic-ray sources within reasonable limits, we must assume that the particles originate inside our galaxy, the Milky Way. According to the best estimate, the Milky Way contains approximately 100 billion stars, most of them confined to a flat volume shaped roughly like a grindstone. The diameter of the galaxy is about 100,000 light years, its thickness about 1,500 light years. In the space between stars there are enormous wandering clouds of gas, mostly hydrogen. If the gas of the clouds were distributed uniformly throughout the galaxy, its density would be about one atom per cubic centimeter. Despite this extreme dilution, the total amount of matter in the clouds is approximately equal to the amount of matter in the stars. In addition to the clouds of gas, there are clouds of dust whose total mass is about one per cent that of the stars.

A question hotly debated among astronomers at present is whether magnetic fields exist in the galaxy. From the theoretical point of view, one would expect the wandering gas clouds to carry with them magnetic fields created at the expense of their kinetic energy. From the experimental side, it is known that the light of distant stars is partly polarized, and it is generally assumed that this polarization occurs when the light passes through dust clouds made of microscopic elongated grains oriented in certain preferred directions. A magnetic field provides a natural explanation for the orientation of the dust particles. However, there are other possible explanations, and we still lack conclusive astronomical evidence for the existence of magnetic fields in interstellar space.

This question is of very great importance in the interpretation of cosmic rays. If there is no magnetic field, cosmic-ray particles travel along straight lines through the Milky Way until they escape or collide with nuclei of hydrogen in interstellar matter. A proton has about one chance in 500 of undergoing a collision on traversing the whole galaxy; the heavier particles have a somewhat greater collision probability. Since the sun is situated in the median plane of the galaxy about 30,000 light years from its center, one would expect that, in the absence of magnetic fields, cosmic rays should reach us at different intensities from different directions. A larger fraction of the cosmic radiation should come from the direction of the center.

Nothing of the sort has been observed: the intensity of cosmic radiation from the different regions of the sky does not vary by more than a fraction of one per cent. This uniformity suggests very strongly that magnetic fields are operating to produce a random distribution of cosmic rays in space. Indeed, the cosmic-ray evidence is perhaps the strongest argument presented so far for the existence of these fields.

One may assume that comparatively strong fields exist inside the gas clouds, while the space between clouds is almost completely field-free. Then a cosmic-ray particle travels a straight path until it strikes a cloud. Once inside the cloud, it describes a complicated orbit and eventually emerges with practically no recollection of the original direction of its motion. According to this model, cosmic rays scatter back and forth between clouds, and it may take a particle produced near the median plane a very long time to reach the border and escape into intergalactic space. Some particles would never do so; they would come to the end of their lives by colliding with atomic nuclei of interstellar hydrogen.

The composition of cosmic rays offers an important clue as to the extent to which cosmic rays may be trapped within the Milky Way. When an alpha particle or a heavier nucleus collides with a nucleus of interstellar hydrogen, it breaks up largely into free neutrons and protons. Free neutrons change into protons with a mean life of about 20 minutes. Thus every collision of an alpha particle or heavier nucleus produces several protons, and these protons have approximately the same velocity as the original particle. We know how many

OBSERVATORY ON MOUNT WRANGELL in southeastern Alaska was established this summer by New York University and the University of Alaska. Its altitude (14,000 feet) and latitude (62 degrees North) make it a unique station for the study of cosmic rays.

alpha particles and heavier nuclei are present in cosmic rays and how far they travel, on the average, before colliding. If we assume that these particles never escape from the galaxy, we can compute the number of protons thus produced. We find that this number is more than enough to account for all of the protons in cosmic rays, even if we assume that no protons are accelerated directly by the mechanism responsible for production of the other components.

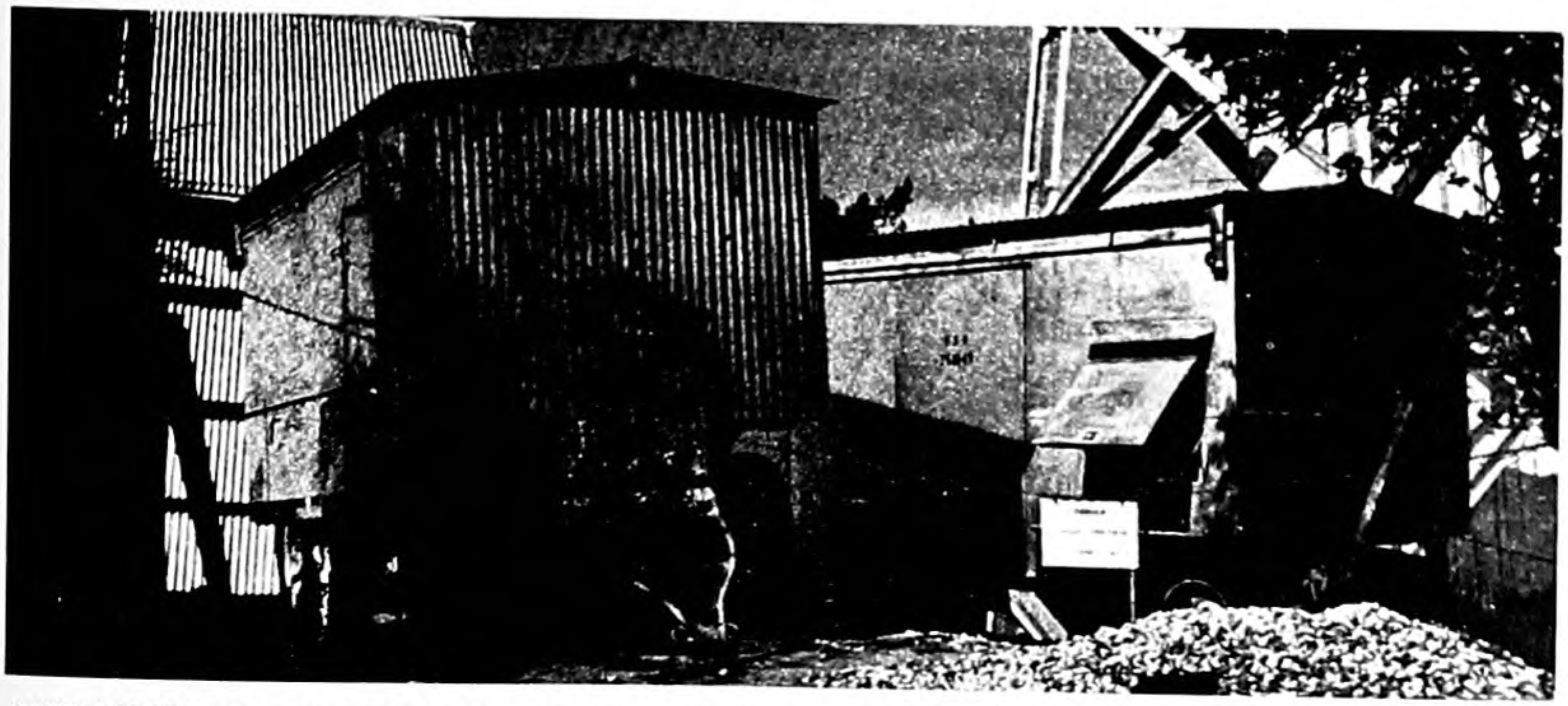
In short, our observations prevent us from assuming either that cosmic rays escape freely from the galaxy (because that is contrary to their distribution) or that they are completely trapped (because that is contrary to their composition). But there seems to be no serious objection to the assumption that they are partly trapped. Their distribution can be accounted for if the mean life of the particles before escape from the galaxy is of the order of several million years, and in that period a large fraction of the heavier nuclei would avoid disintegrating collisions (on the average an alpha particle, for example, collides with another particle once in 170 million years).

It is reasonable to suppose that in most stars particles are accelerated by processes similar to those which take place in the sun. If all cosmic rays originate from stars, the average star in the galaxy must produce several million times more cosmic radiation than the sun, and some stars must eject particles of very much higher energy, to account for the radiation we receive. We know, as a matter of fact, that the sun is a comparatively inactive star. It is possible that the bulk of the radiation may come from a few types of stars with special characteristics. Double stars, whose two partners possess strong magnetic moments and revolve rapidly around a common center of gravity, and variable stars, whose exceedingly strong magnetic fields appear to reverse direction in a period of a few days, have been singled out as important cosmic-ray sources. But the most prominent candidates of all are the supernovae.

Supernovae are gigantic explosions. One occurs every few hundred years in our galaxy. The explosion releases an amount of energy perhaps equivalent to the total mass of the sun. According to some theories, a star bursts into a supernova when it has exhausted its supply of hydrogen (the fuel that keeps the star hot) and its rotational velocity has dropped below a critical level. At this point gravitational attraction, no longer balanced by centrifugal force or by internal pressure, causes the star to collapse until its core is just one enormous lump of nuclear matter. The star blows up in a tremendous explosion whose products are mainly nuclei of the heavy stable elements. It is possible that cosmic rays are secondary products of this outburst. In the initial phase of the explosion lumps of nuclear matter much heavier than the nuclei of any of the ordinary elements may be thrown into space with velocities approaching the speed of light. These fragments would be highly unstable and would disintegrate, ejecting protons, alpha particles and other nuclei. The ejected particles, with energies further increased by electrostatic repulsion of the highly charged fragment from which they originated, would form the cosmic radiation. Even if cosmic rays were produced entirely in instantaneous bursts, several hundred years apart, we should still expect the observed cosmic ray intensity to be constant, because the particles presumably circulate in the Milky Way for millions of years.

Lyman Spitzer of Princeton University has suggested another idea. Dust particles in the neighborhood of a supernova, accelerated by the pressure of the tremendous light flash that accompanies the outburst, might break up into cosmic rays upon colliding with nuclei of hydrogen in space. This process does not account very well for the production of particles with energies above several billion electron volts.

All the theories considered so far have one feature in common: they assume that cosmic-ray particles attain their full energy in a very short time, through some violent process occurring either in a star or in the immediate neighborhood of a star. One must also consider the alternative possibility that these particles are not energetic to start with but acquire their high energy gradually by the action of extended electromagnetic fields as they travel through space. An acceleration mechanism of this kind has been suggested by Enrico Fermi, who points out that energy is exchanged whenever cosmic-ray particles collide with gas clouds. When a particle collides with a cloud coming toward it, it bounces off



OBSERVATORY ON MOUNT WILSON in southern California is used to study cosmic rays by physicists from the California Institute of Technology. It is best known for the work done there on the V-particles, elementary constituents of matter heavier than protons.

with increased energy; when it collides with a receding cloud, it loses energy. On the average, a particle obtains a small net gain of energy, largely because the number of head-on collisions is slightly greater than the number of overtaking collisions. This energy gain is proportional to the energy of the particle at any given moment. Therefore the particle energy increases exponentially with time, like a savings account.

The Fermi mechanism is effective only if the average energy gain per collision with a cloud is greater than the energy loss by ionization in interstellar matter between collisions. Since the energy gain increases and the energy loss decreases with increasing particle energy, this requirement sets a lower limit to the energy of the particles that can be further accelerated. In other words, particles must be shot into space with more than a certain energy, which is different for different kinds of particles. Fermi estimated that the minimum injection energy is about 200 million electron volts for protons, about one billion for alpha particles, and about 20 billion for oxygen nuclei and increasingly larger for the heavier nuclei.

The large value of the injection energy required for the heavier components makes it difficult to understand how these can be present with any appreciable intensity. Moreover, alpha particles and the heavier nuclei collide with interstellar hydrogen much more frequently than protons do. Therefore they have a smaller chance of gaining large amounts of energy through the suggested acceleration mechanisms. Thus Fermi's original theory, which assumes a complete trapping of the particles in the galaxy, does not account for the fact that the energy distribution of alpha particles and heavier nuclei in the cosmic radiation is similar to that of protons.

One may possibly find a way out of these difficulties by the assumption, already made earlier in this article, that the cosmic-ray particles are only partly trapped, so that their mean life in the galaxy is determined primarily by their probability of escape rather than by their probability of collision. The mean life of protons, alpha particles and heavier nuclei is then approximately the same, and their energy distribution is likewise approximately the same.

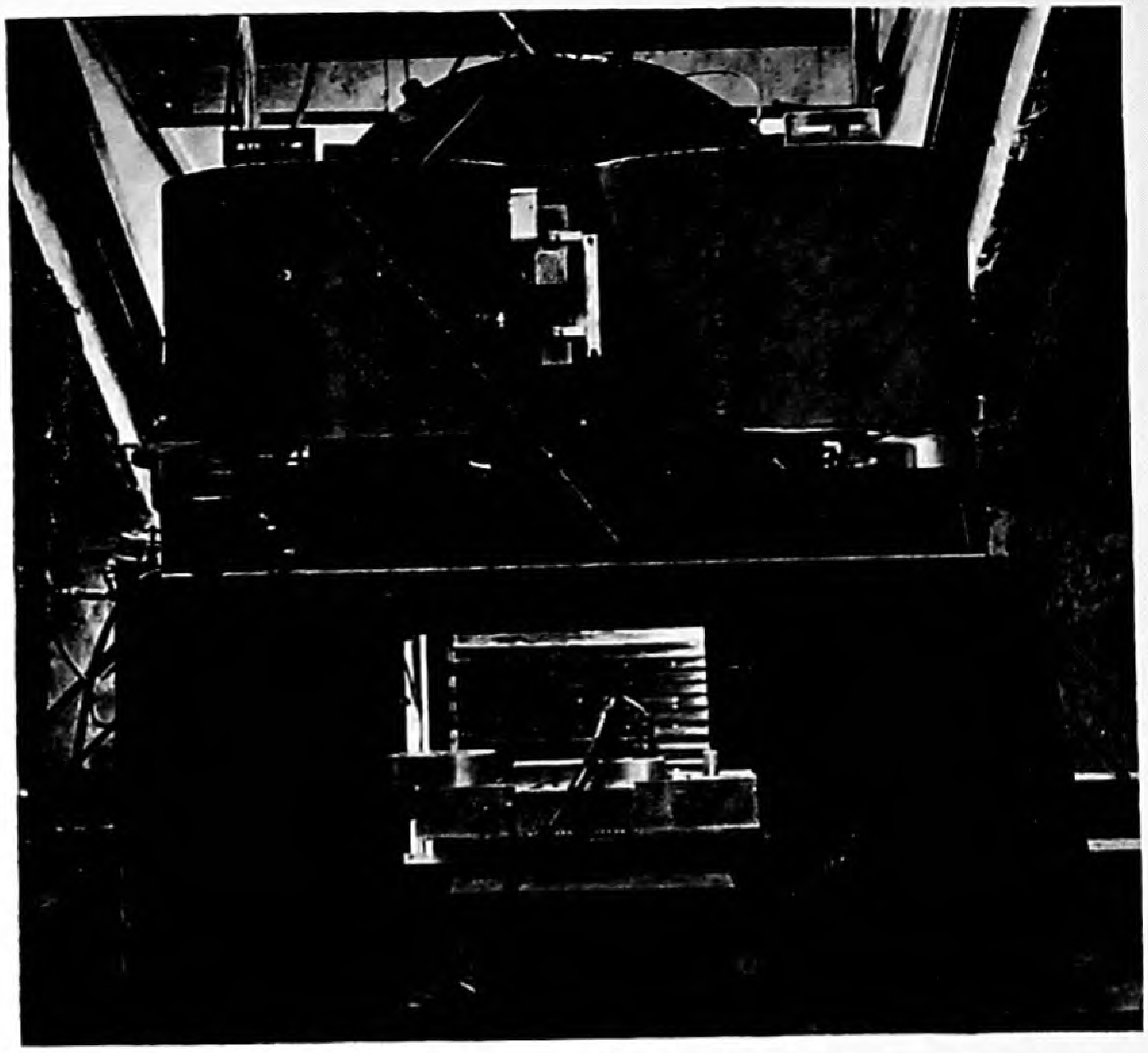
The energy spectrum of particles accelerated by Fermi's mechanism depends on the product of the mean life and the average energy gain per unit of time. Since, by assuming only partial trapping, we have decreased the estimated lifetime, we must correspondingly increase the estimate of the energy gain. This decreases the required injection energies and thus makes the Fermi mechanism more plausible for the heavy nuclei.

At present no hypothesis about the origin of cosmic rays is unequivocally supported by theory or experiment. What is worse, few of the many hypotheses that have been put forward can be definitely disproved. I favor the following general picture: The sun and the other stars eject protons and nuclei of heavier elements at energies comparable to those now attained by man-made accelerators. They do so by an electromagnetic process or a combination of such processes. Some stars may be unusually strong sources of the particles. The particles then diffuse through interstellar space, being trapped within the galaxy by irregularly distributed magnetic fields for an average of several million years. During their random motion they undergo further acceleration through some mechanism of the kind suggested by Fermi.

One can think of several investigations to test this thoroughly tentative picture. The fact that cosmic radiation contains no protons with less than about one billion electron volts may be an important clue. It would be interesting to know whether low-energy protons are also absent from the solar radiation that reaches the earth after flares. The lack of high-energy electrons and photons in cosmic rays needs investigation. Collisions of cosmic-ray particles with interstellar hydrogen produce pi mesons, which then decay into mu mesons, which then decay into electrons. They also produce neutral mesons, which decay into photons. Thus some electrons and photons should be present in the cosmic radiation, even if these particles are not produced at the source. They may have escaped detection thus far only because their number is small.

The magnetic field of the galaxy cannot effectively trap particles beyond a certain energy. Thus there should be a high-energy cut-off in the energy spectrum of cosmic rays. If this cut-off could be determined experimentally, it would afford valuable evidence on the strength of the magnetic fields in the galaxy.

There is little doubt that cosmic-ray research eventually will contribute as much to our knowledge of the universe as it has contributed to our knowledge about particles and nuclear forces.



CLOUD CHAMBER AND MAGNET are employed in the investigation of cosmic rays at the University of California. In chamber at bottom are a series of horizontal lead plates.

The Author

BRUNO ROSSI became "fascinated by the mystery of this extraordinary phenomenon" through reading an article about it in 1930. He at once began building his first Geiger-Müller counter, and has been building them ever since. Rossi was born in Venice in 1905 and was educated at the Universities of Padua and Bologna. He received his Ph.D. in physics in 1937. He taught at Florence and then at Padua until he was forced out by the Fascist government in 1938. He then spent a year in Denmark and England and came to the U.S. in 1939 at the invitation of Arthur H. Compton of the University of Chicago. Rossi was an associate professor of physics at Cornell University from 1940 to 1946, with three years out for work at Los Alamos. Since 1946 he has been professor of physics at the Massachusetts Institute of Technology.

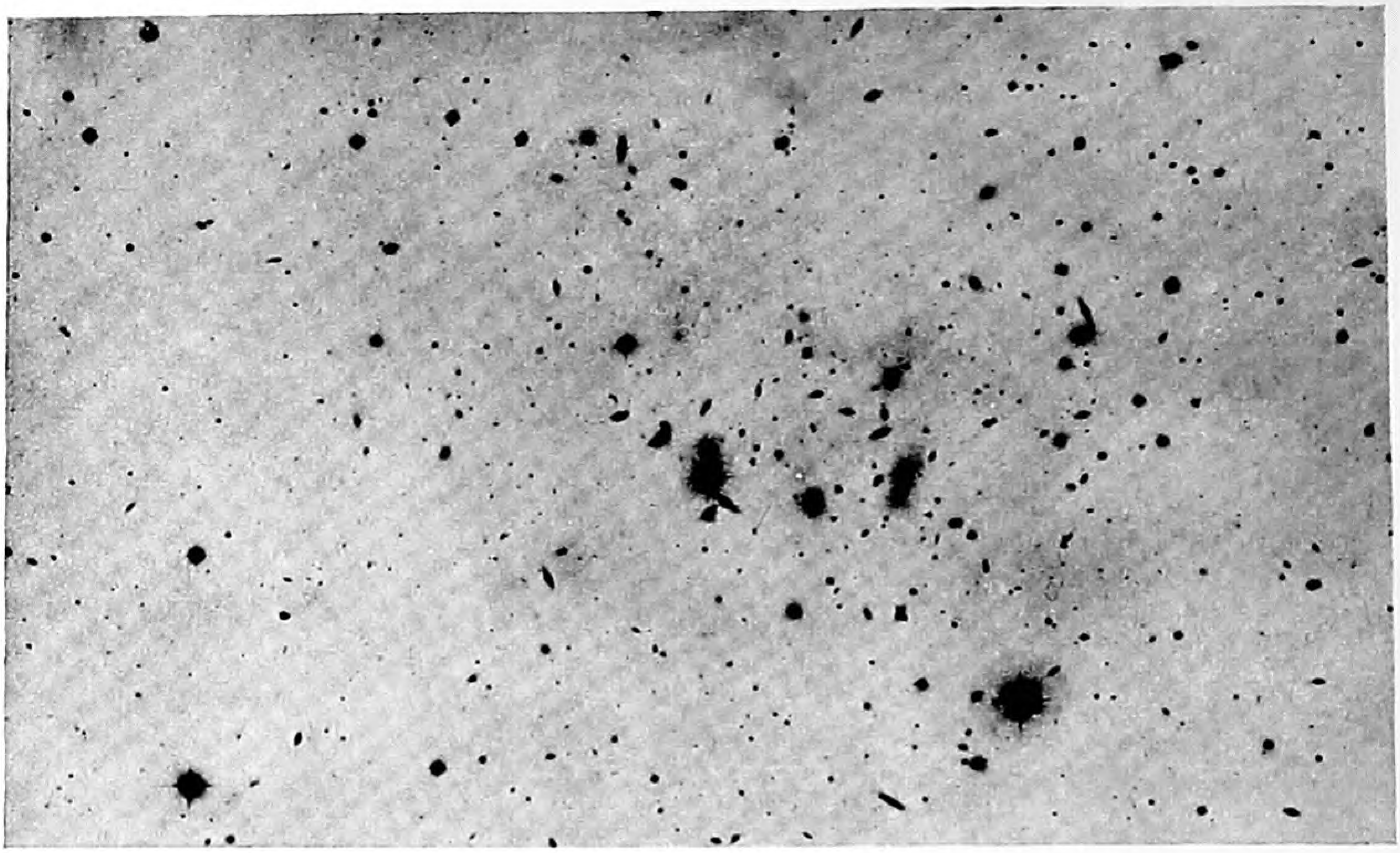
Bibliography

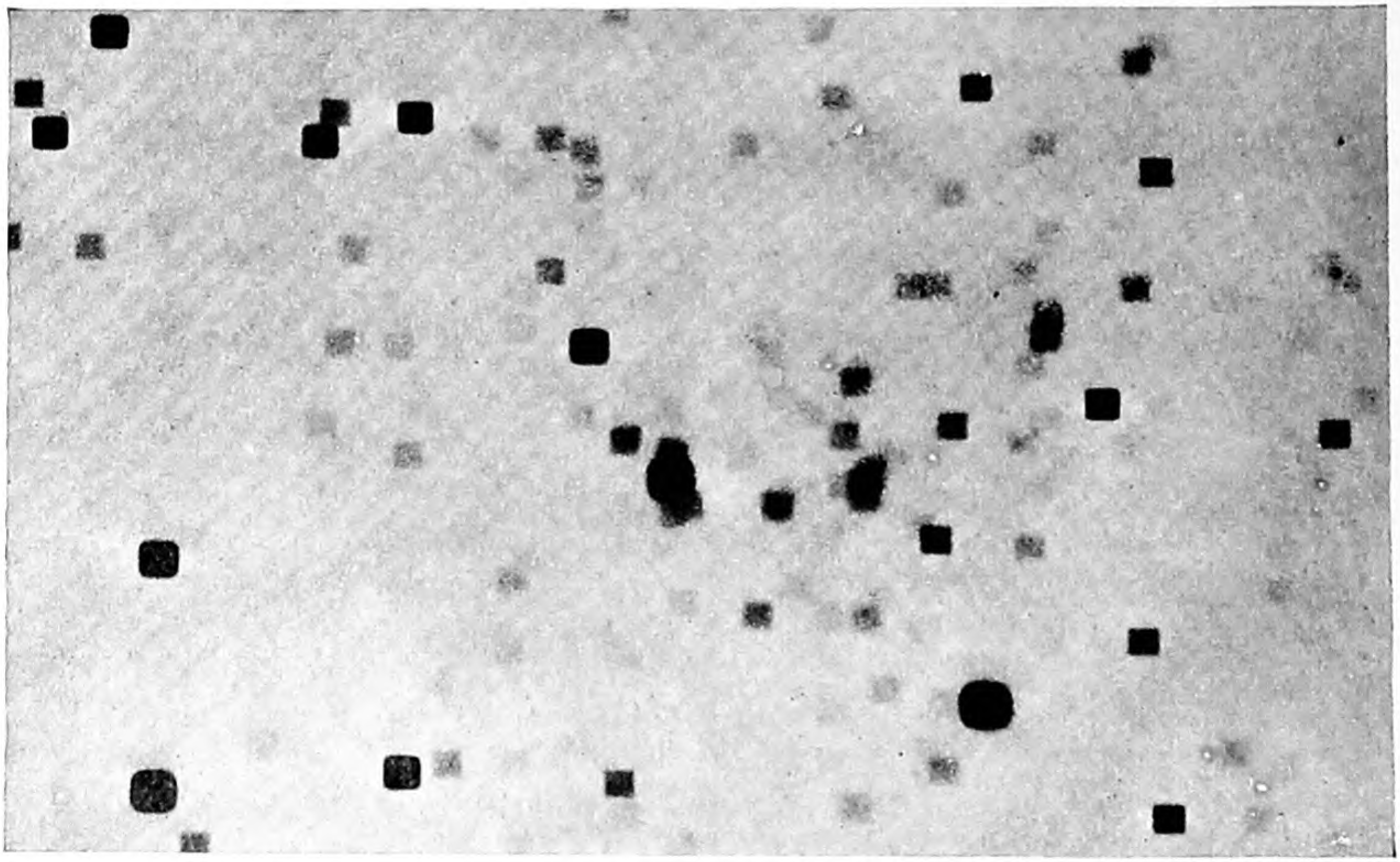
HIGH ENERGY PARTICLES. Bruno Rossi. Prentice-Hall, Inc., 1952.

THE COSMIC RAYS. Louis Leprince-Ringuet. Prentice-Hall, Inc., 1950.

PROGRESS IN COSMIC RAY PHYSICS. Edited by J. G. Wilson. Interscience Publishers, Inc., 1951.

COSMIC RAYS. L. Jánossy. Oxford University Press, 1950.





BRIGHTNESS OF GALAXIES may be measured with the help of the jiggle camera (*see photograph on the opposite page*). At the top is a negative print of a 200-inch telescope photograph showing nearby stars and a cluster of galaxies in Corona Borealis. Although the stars have made the brightest images in the photograph, they are essentially point sources of light. The galaxies, on the other hand, are extended sources of light. To measure the brightness of a galaxy by comparing it with the known brightness of a star, the two images must be made the same. This is done by smearing the images as shown at bottom in a jiggle-camera photograph of the same area.

SCIENTIFIC AMERICAN September 1956, Vol. 195, No. 3, pp. 170-182 OFFPRINT 240

THE RED-SHIFT

by Allan R. Sandage

The redness, and presumably the speed of recession, of most galaxies increases regularly with distance. The most distant galaxies observed appear to depart from this law, a fact of deep meaning for cosmology.

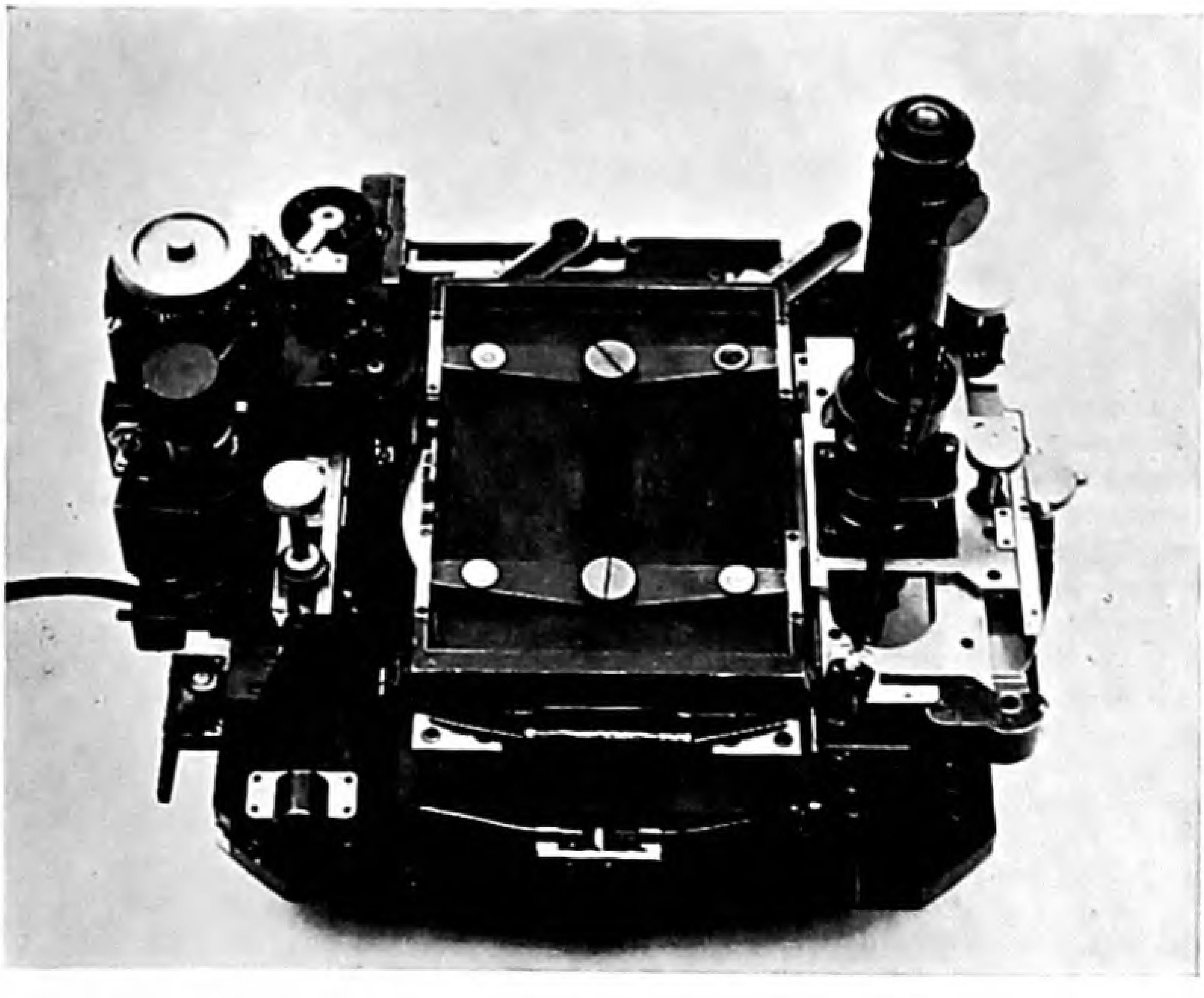
In the nature of things it is a delicate undertaking to try to discern the general structure and features of a universe which stretches out farther than we can see. For more than a quarter of a century both the theoreticians and the observers of the cosmos have been making exciting discoveries, but the points of contact between the discoveries have been few. The predictions of the theorists, deduced from the most general laws of physics, are not easy to test against the real world—or rather, the small portion of the real world that we can observe. There is, however, one solid meeting ground between the theories and the observations, and that is the apparent expansion of the universe. Other aspects of the universe may be interpreted in different ways to fit different theories, but concerning the expansion the rival theories make unambiguous predictions on which they will stand or fall. There is now hope that red-shift measurements of the universe's expansion with the 200-inch telescope on Palomar Mountain will soon make it possible to decide, among other things, whether we live in an evolving or a steady-state universe.

Let us begin by considering just what issue the measurements seek to decide, as between the two principal competing theories in modern cosmology. The steady-state theory says that the universe has been expanding at a constant rate throughout an infinity of time. The evolutionary theory, in contrast, implies that the expansion of the universe is steadily slowing down. If the universe began with an explosion from a superdense state, its rate of expansion was greatest at the beginning and has been slowing ever since because of the opposing gravitational attraction of its matter, which acts as a brake on the expansion—much as an anchored elastic string attached to a golf ball would act as a brake on the flight of the ball.

Now in principle we can decide whether the rate of expansion has changed or not simply by measuring the speed of expansion at different times in the universe's history. And the 200-inch telescope permits us to do this. It covers a range of about two billion years in time. We see the nearest galaxies as they were only a few million years ago, while the light from the most distant galaxies takes so long to reach us that we see them at a stage in the universe's history going back to one or two billion years ago. If the explosion theory is correct, the universe should have been expanding at an appreciably faster rate then than it is now. Since the light we are receiving from the distant galaxies is a flashback to that earlier time, its red-shift should show them receding from us faster than if the rate of expansion had remained constant.

The red-shift is so basic a tool for testing our notions about the universe that it is worthwhile to review how it was discovered and how it is used.

An astronomer cannot perform experi-



JIGGLE CAMERA smears the images by moving the photographic plate in a rectangle during exposure. It is mounted in the prime-focus cage at the upper end of the 200-inch.

ments on the objects of his study, or even examine them at first hand. All his information rides on beams of light from outer space. By sufficiently ingenious instruments and equally ingenious interpretation (we hope), he may translate this light into information about the temperatures, sizes, structures and motions of the celestial bodies. It was in 1888 that a German astronomer, H. C. Vogel, first demonstrated that the spectra of stars could give information about motions which could not otherwise be detected. He discovered the Doppler effect in starlight.

The Doppler effect, as every physics student knows, is a change in wavelength observable when the source of radiation (sound, light, etc.) is in motion. If it is moving toward the observer, the wavelength is shortened; if away, the waves are lengthened. In the case of a star moving away from us, the whole spectrum of its light is shifted toward the red, or long-wave, end.

This spectrum, made by means of a prism or diffraction grating which spreads the light out into a band of its component colors, is usually not continuous. Certain wavelengths of the light are absorbed by atoms in the star's atmosphere. For example, most stars show strong absorption, by calcium atoms, at the wavelengths of 3933.664 and 3968.-470 Angstrom units. (An Angstrom unit is a hundred-millionth of a centimeter.) The absorption is signaled by dark lines in the spectrum, known in this case as the K and H lines of calcium. Now if a star is moving away from us, these lines will be displaced toward the red end of the spectrum. In the spectrum of the star known as Delta Leporis, for in-

VIRGO

22 MILLION LIGHT-YEARS

1,200 KILOMETERS PER SECOND

CORONA BOREALIS

400 MILLION LIGHT-YEARS

21,500 KILOMETERS PER SECOND

BOOTES

700 MILLION LIGHT-YEARS

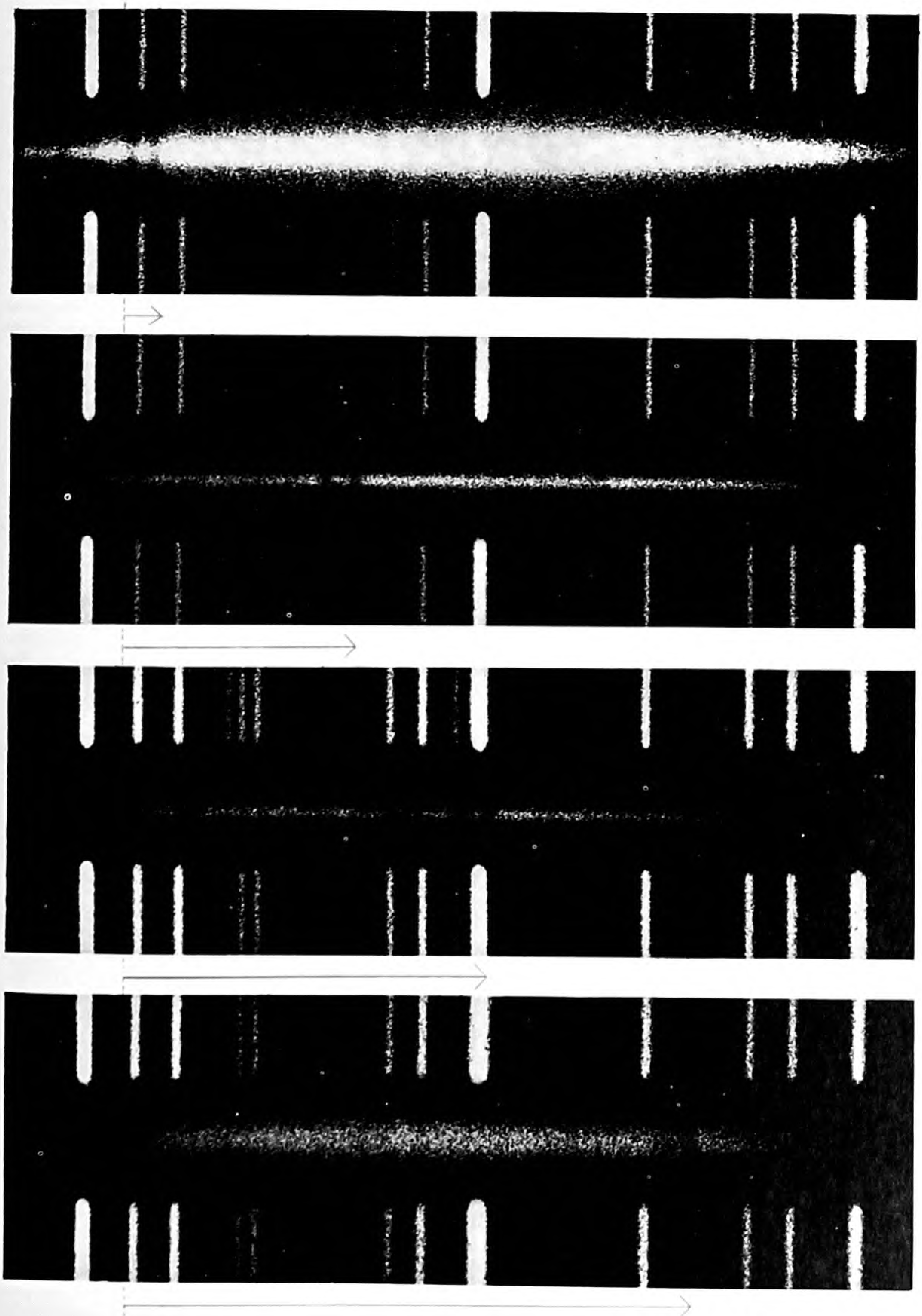
39,300 KILOMETERS PER SECOND

HYDRA

1.1 BILLION LIGHT-YEARS

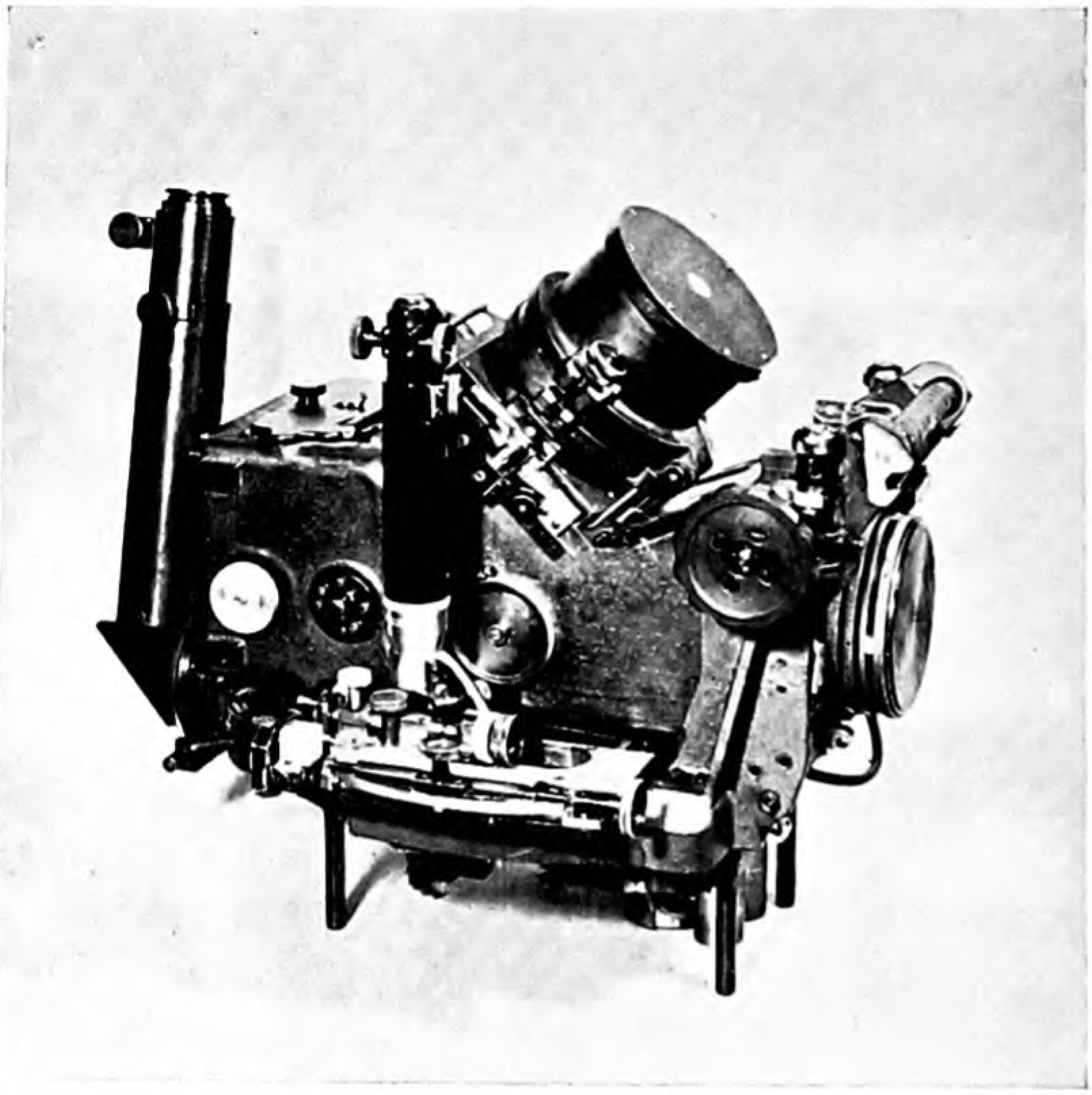
60,900 KILOMETERS PER SECOND

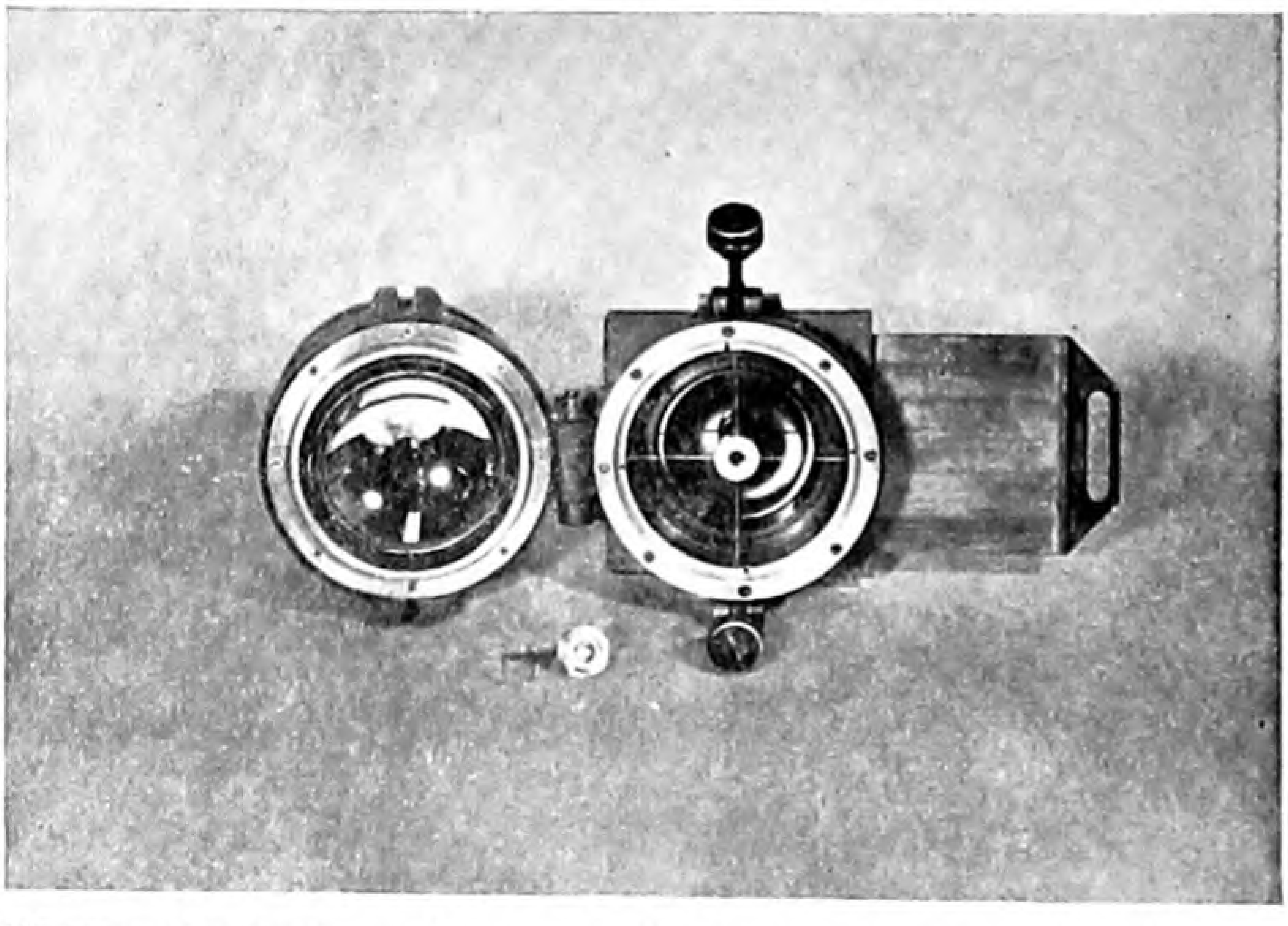
RED-SHIFT of four galaxies on this page is depicted in the spectra on the opposite page. The galaxies are centered in the photographs. The spectra are the bright horizontal streaks tapered to the left and right. Above and below each spectrum are comparison lines from the spectrum of iron. Near the left end of the spectrum at the top of the page are two dark vertical lines: the K and H lines of calcium. If the galaxy did not exhibit the red-shift, these lines would be in the position of the broken line running vertically down the page. The amount of their shift toward the red, or right, end of the spectrum is indicated by the short arrow to the right of the broken line. The larger shift of the K and H lines of the three fainter galaxies is indicated by the longer arrows below their spectra. The constellation, approximate distance and velocity of recession of each galaxy is at left of its photograph.



stance, the K line of calcium is displaced 1.298 Angstroms toward the red. Assuming the displacement is due to the Doppler effect, it is a simple matter to calculate the velocity of the star's receding motion. Dividing the amount of the displacement by the normal wavelength at rest, and multiplying by the speed of light (300,000 kilometers per second) we get the speed of the star—in this case 99 kilometers per second. The calculation on the basis of displacement of the H line gives the same figure.

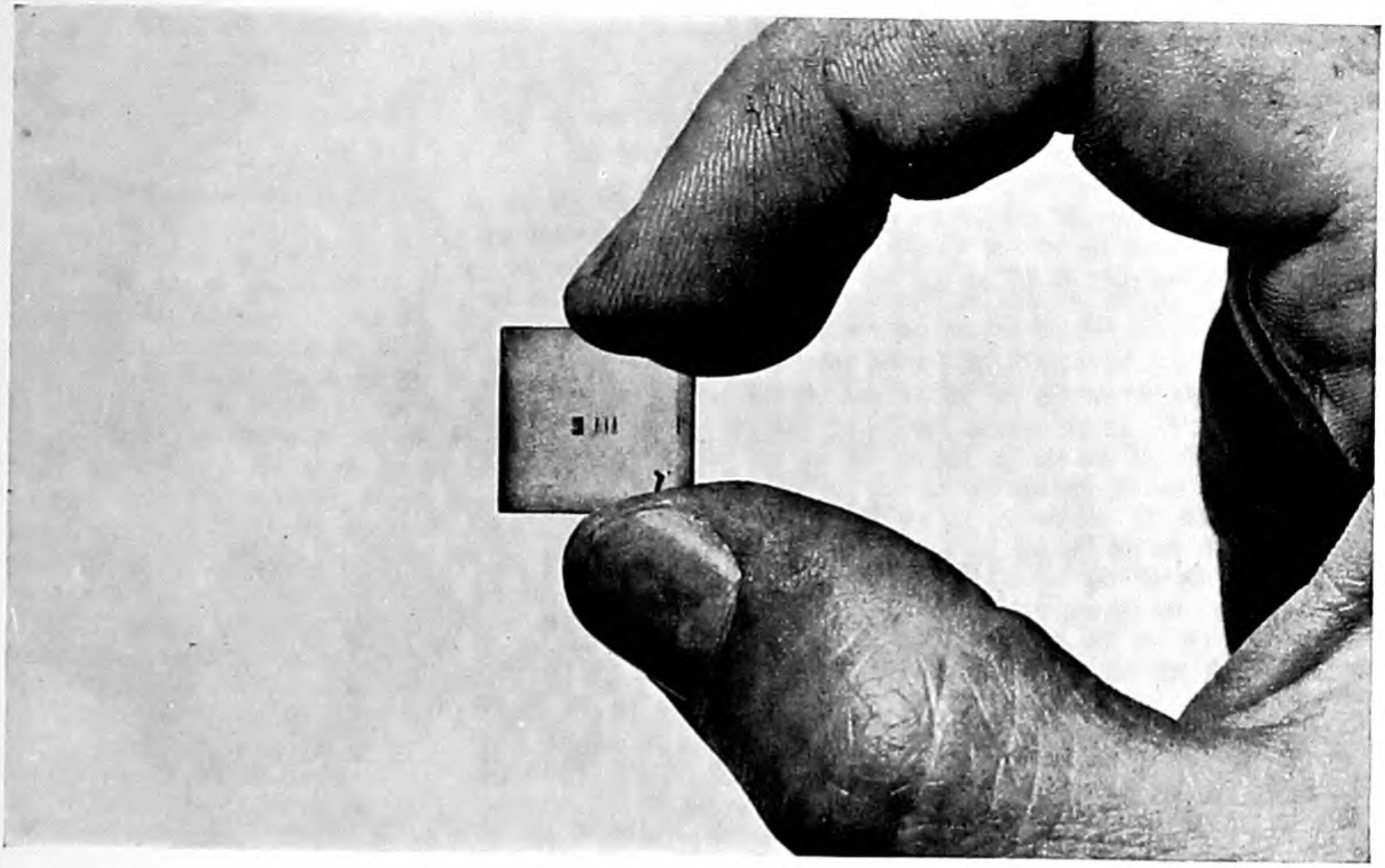
Equipped with this powerful tool, many of the large observatories in the world spent a major part of their time during the early part of this century measuring the velocities of receding and approaching stars in our galaxy. At first it was a work of pure curiosity, no one suspecting that it might have any bearing on cosmological theories. But in the 1920s V. M. Slipher of the Lowell Observatory made a discovery which was to lead to a completely new picture of the universe. His measurements of red-shifts of a number of "nebulae" then thought to lie in our galaxy showed that they were all receding from us at phenomenal speeds—up to 1,800 kilometers per second. Edwin P. Hubble at Mount Wilson soon established that the "nebulae" were systems of stars, and he went on to measure their distances. The method he used was the one developed by Harlow Shapley, employing Cepheid variable stars as the yardstick. Shapley had found a way to measure the intrinsic brightness of these stars, and therefore their distance could be estimated from their apparent brightness by means of the rule that the intensity of light falls off as the square of the distance. Hubble observed that the galaxies nearest our own system, including the Great Nebula in Andromeda, contained Cepheid variables, and when he computed their distances he came out with the then astounding figure of about one million light-years! He next tackled the problem of finding the distances of Slipher's nebulae. Since variable stars could not be detected in them, he used their brightest stars as distance indicators instead. He found that these nebulae were at distances ranging up to 20 million light-years from us, and what was more remarkable, their velocities increased in strict proportion to their distances!





SPECTRA ARE MADE with the spectrograph at the top, which is mounted at the prime focus of the 200-inch telescope. Inside the spectrograph the converging rays of the 200-inch mirror are made parallel by a concave mirror. The light is then dispersed by a diffraction grating. At the bottom is a Schmidt camera used to photograph the spectrum. It has an optical path of solid glass and a speed of $f/.48$. The plateholder and plate are below the camera.

Hubble made the daring conjecture that the universe as a whole was expanding. He predicted that more remote galaxies would show larger red-shifts, still in proportion to their distance. To test Hubble's speculation, Milton L. Humason began a long-range program of spectral analysis of more distant galaxies with the 100-inch telescope on Mount Wilson. In these faint galaxies it was no longer possible to distinguish even the bright stars, and so the relative brightness of the galaxy as a whole had to be taken as the measure of distance. That is, a galaxy one fourth as bright as another was assumed to be twice as far away. Hubble reasoned that while individual galaxies might deviate from this rule, statistically the population of galaxies as a whole would follow it. The prin-



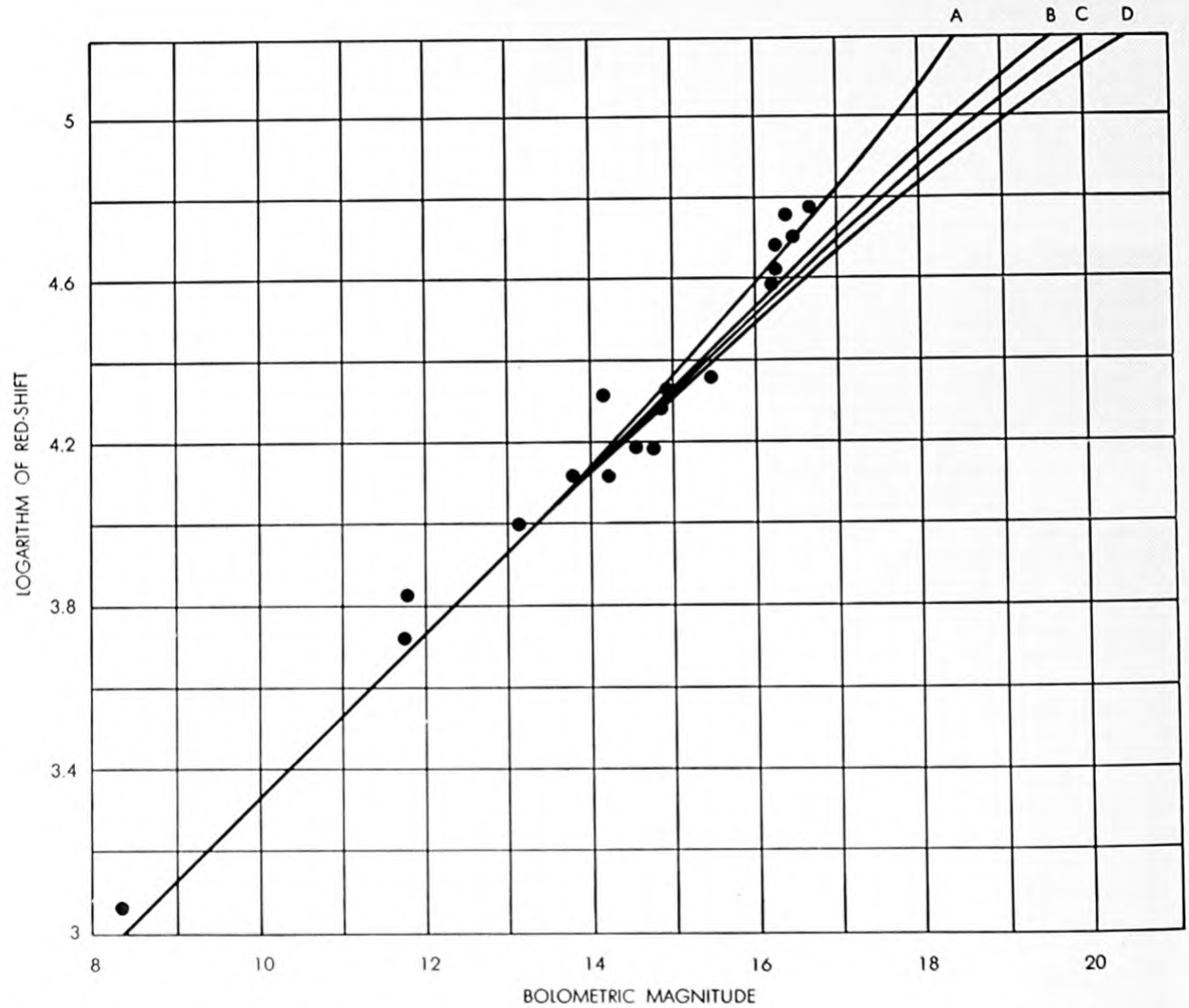
ACTUAL SIZE of the red-shift spectrum is indicated by the photograph at the top of the page. The glass photographic plate is 15 millimeters on an edge. The spectrum is 5 mm. long. At bottom Milton L. Humason examines a spectrum with a low-power microscope.

ciple is still the basis of distance determinations today.

Humason laboriously photographed spectra of galaxies, and Hubble measured their apparent brightness, from 1928 to 1936, when they reached the limit of the 100-inch telescope. The history of the red-shift program in those years is a story of extreme skill and patience at the telescope and of steady improvement in instrumentation. It was a long and difficult task to photograph spectra then; the prisms used required long exposures, and it took 10 nights or more to obtain a spectrum which with modern equipment can be recorded in less than an hour today. The improvement in equipment includes not only the 200-inch telescope but also diffraction gratings, faster cameras and a vast improvement in the sensitivity of photographic plates, thanks to the Eastman Kodak Company. Astronomers the world over, and cosmology, owe a large debt to the Eastman research laboratories.

Humason's first really big red-shift came early in 1928, when he got a spectrum of a galaxy called NGC 7619. Hubble had predicted that its velocity should be slightly less than 4,000 kilometers per second: Humason found it to be 3,800. By 1936, at the limit of the 100-inch telescope's reach, they had arrived at a cluster of galaxies, called Ursa Major No. 2, which showed a velocity of 40,000 kilometers per second. All the way out to that range of more than half a billion light-years the velocity of galaxies increased in direct proportion to the distance. In a sense this was disappointing, because the various cosmological theories predicted that some change in this relation should begin to appear when the observations had been pushed far enough. Further exploration into the distances of space had to await the completion of the 200-inch Hale telescope on Palomar Mountain.

In 1951 the red-shift program was resumed, with a new spectrograph of great speed and versatility placed in the big telescope's prime focus cage, where the observer rides with his instruments. The spectrograph has to be of very compact design to fit into the cramped space of the cage. The photographic plate itself, mounted in the middle of a complex



EIGHTEEN FAINTEST CLUSTERS of galaxies yet measured are plotted for their red-shift (or speed of recession) and apparent magnitude (or distance). Line C is a universe expanding forever at the same rate. Line D is a steady-state universe. If the line falls to the left of C, the expansion must slow down. If it falls between C and B, the universe is open and infinite. If it falls to the left of B, the universe is closed and finite. If it falls on B, it is Euclidean and infinite. A is the trend suggested by the six faintest clusters.

optical arrangement, is only 15 millimeters (about half an inch) on a side. The cutting and handling of such small pieces of glass in complete darkness (to avoid exposure of the plate) is a tricky business. The spectrum recorded on the plate is a tiny strip only a fifth of an inch long, but it is long enough to measure red-shifts to an accuracy of better than one half of 1 per cent.

The most distant photographable galaxies are so faint that they are not visible to the eye through the telescope: they can be recorded only by extended exposure of the plate. The observer guiding the telescope must position the slit of the spectrograph by reference to guide stars within the same field as the distant object. Another great difficulty in recording the red-shift of extremely distant galaxies arises from the magnitude of the shift. The displacement of the calcium dark lines toward the red is so large that the lines move clean off the sensitive range of blue photographic plates, which astronomers like to use because of their speed. So slow panchromatic plates must be used, and Humason has been forced to return to exposure times as long as 30 hours or more.

The other part of the program—measuring the distances of the galaxies—also has been helped by improvements in technique. For measurement of their brightness the Mount Wilson telescopes employ photomultiplier tubes, which amplify the light energy by electronic means. Such equipment was not available for the 200-inch telescope when the present program began. Instead the intensity of the light from very faint galaxies was measured by a tricky method which compares it with that of stars of known magnitude. No direct comparison can be made, of course, between the picture of a star and that of a galaxy or cluster of galaxies, because the star is a point source of light while a galactic system is a spread-out image. To make the images comparable, a region of the sky is photographed with a "jiggle" camera which moves the plate around so that the images of stars and of galaxies are smeared out in squares [*see photograph at bottom of page 330*]. They can then be compared as to brightness—just as one may use color cards to find a match to the color of a room.

Humason has now measured red-shifts of remote clusters of galaxies with recession velocities up to 60,000 kilometers per second. What do they show? Is the velocity still increasing in strict proportion to the distance?

The information about 18 of the faintest measured clusters is given in the accompanying chart [*see chart on page 336*]. The velocities are plotted against their apparent brightness, or estimated distances. If velocity increases in direct proportion to the distance, the observed velocity-distance relation should be "linear" (*i.e.*, follow a straight line). But as the chart shows, the very faintest clusters have begun to depart from that line. These clusters, about a billion light-years away, are moving *faster* (by about 10,000 kilometers per second) than in direct proportion to their apparent distance. In other words, the data would be interpreted to mean that a billion years ago the universe was expanding faster than it is now. If the measurements and the interpretation are correct, this suggests that we live in an evolving rather than in a steady-state universe.

The observed change in the curve buys us much more information. To begin with, it tells us something about the mean density of matter in the universe. The rate at which the expansion of the universe is slowing down (if it is) depends on the mean density of its matter: the higher the density, the greater the braking effect. The amount of departure from linearity indicated by the measurements thus far calls for a mean density of about 3×10^{-28} grams of matter per cubic centimeter (about one hydrogen atom per five quarts of space). Now this amounts to about 300 times the total mass of the matter estimated to be contained in galaxies: that figure comes out to a mean density of only 10^{-30} grams per cubic centimeter. If our present tentative value for the slowdown of the expansion should be confirmed, we would have to conclude that either the current estimates of the masses of the galaxies are wrong or that there is a great deal of matter, so far undetected, in intergalactic space. Matter in the form of neutral hydrogen (*i.e.*, normal hydrogen atoms consisting of a proton and an electron) might be present in space and still have escaped detection until now because it is not luminous. The giant radio telescopes now under construction or on the drawing boards perhaps will detect the hydrogen, if it exists in the postulated quantities.

Once we know the rate at which expansion of the universe is slowing down, it becomes possible to determine not only the mean density of matter but also the geometry of space—that is, its curvature. Models of the evolving universe take three forms: the Euclidean case, in which space is flat, open and infinite; a curved universe which is closed and finite, like the surface of a sphere; and a curved universe which is open and infinite, like the surface of a saddle. In the accompanying velocity-distance chart [*see page 336*], curves to left of C represent evolving models, and curve D represents the steady-state model. If the curve of the velocity-distance relation lies between C and B, the universe is open and infinite. Line B is the Euclidean case of flat space. If the curve is left of B, the universe is closed and finite, the radius of its curvature decreasing to the left.

According to our present observations, the actual relation follows a curve left of B (curve A on the chart). Although our data are still crude and inconclusive, they do suggest that the steady-state model does not fit the real world, and that we live in a closed, evolving universe.

Humason has gone beyond 60,000 kilometers per second and attempted to measure the red-shifts of two faint clusters whose predicted velocity is more than 100,000 kilometers per second. So far these efforts have not yielded reliable results, but he is continuing them. These two remote clusters may well hold the key to the structure of the universe. We stand a chance of finding the answer to the cosmological problem. The red-shift program will continue toward this goal.

If the expansion of the universe is decelerating at the rate our present data suggest, the expansion will eventually stop and contraction will begin. If it returns to a superdense state and explodes again, then in the next cycle of oscillation, some 15 billion years hence, we may all find ourselves again pursuing our present tasks.

Although no final answers have yet emerged, big steps have been taken since 1928 toward the solution to the cosmological problem, and there is hope that it may now be within our grasp. The situation has nowhere been better expressed than in Hubble's last paper:

"For I can end as I began. From our home on the earth we look out into the distances and strive to imagine the sort of world into which we are born. Today we have reached far out into space. Our immediate neighborhood we know rather intimately. But with increasing distance our knowledge fades . . . until at the last dim horizon we search among ghostly errors of observations for landmarks that are scarcely more substantial. The search will continue. The urge is older than history. It is not satisfied and it will not be suppressed."

The Author

ALLAN R. SANDAGE is assistant astronomer at the Mount Wilson and Palomar Observatories. He attended Miami University in Ohio, transferred "by the grace of Uncle Sam" to the Navy in 1945 and completed his undergraduate work at the University of Illinois in 1948. He acquired his doctorate in astronomy at the California Institute of Technology in 1953, having meanwhile joined the staff of Mount Wilson and Palomar. Few astronomers have set out on their professional careers so early in life. "As I recall, it was *Buck Rogers in the 25th Century* that steered me into astronomy at the unknowing age of 10. This unfortunate interest (from our neighbors' point of view) took the form of dragging a telescope, tables and other observing paraphernalia into our back yard at three in the morning to look at meteor showers and the like. For some reason the neighbors failed to understand the importance of such operations, and I failed to understand their need for sleep. Today, 20 years later, I sit sleepily at work on Palomar Mountain and wonder how young boys of 10 can take sleep so casually. From two to four A.M. *all* professional astronomers are sleepy. If questioned in this interval, most of them would express serious doubt as to whether astronomy was worth it all. (The doubts always disappear in the morning.) From 1951 to 1953 I was an assistant to the late Edwin P. Hubble. His principal interest was observational cosmology, and a bit of his enthusiasm rubbed off on all who knew him. The associations with Hubble, Milton L. Humason and Walter Baade have been the high points in my scientific career. At present my chief interest is in a slightly different field: that of the observational approach to stellar evolution. This is not so far afield from cosmology as might be supposed, since each discipline is an attempt to rewrite parts of the first book of Genesis. My principal outside activities are long back-pack trips in the California Sierra, fishing in both fresh water and surf, and a recently developed passion for horseback riding."

Bibliography

THE EXPANSION OF THE UNIVERSE. Paul Couderc. Faber & Faber, Ltd., 1952.

THE REALM OF THE NEBULAE. Edwin Hubble. Yale University Press, 1936.

RED-SHIFTS AND MAGNITUDES OF EXTRAGALACTIC NEBULAE. M. L. Humason, N. U. Mayall and A. R. Sandage in *The Astronomical Journal*, Vol. 61, No. 3, pages 97-162; April, 1956.

SCIENTIFIC AMERICAN September 1953, Vol. 189, No. 3, pp. 52-57 OFFPRINT 241

WHAT IS MATTER?

by Erwin Schrödinger

How matter can be both discrete particles and continuous wave patterns of energy is a great dilemma and a fundamental question in science.

Fifty years ago science seemed on the road to a clear-cut answer to the ancient question which is the title of this article. It looked as if matter would be reduced at last to its ultimate building blocks—to certain submicroscopic but nevertheless tangible and measurable particles. But it proved to be less simple than that. Today a physicist no longer can distinguish significantly between matter and something else. We no longer contrast matter with forces or fields of force as different entities; we know now that these concepts must be merged. It is true that we speak of "empty" space (*i.e.*, space free of matter), but space is never really empty, because even in the remotest voids of the universe there is always starlight—and *that* is matter. Besides, space is filled with gravitational fields, and according to Einstein gravity and inertia cannot very well be separated.

Thus the subject of this article is in fact the total picture of space-time reality as envisaged by physics. We have to admit that our conception of material reality today is more wavering and uncertain than it has been for a long time. We know a great many interesting details, learn new ones every week. But to construct a clear, easily comprehensible picture on which all physicists would agree—that is simply impossible. Physics stands at a grave crisis of ideas. In the face of this crisis, many maintain that no objective picture of reality is possible. However, the optimists among us (of whom I consider myself one) look upon this view as a philosophical extravagance born of despair. We hope that the present fluctuations of thinking are only indications of an upheaval of old beliefs which in the end will lead to something better than the mess of formulas which today surrounds our subject.

Since the picture of matter that I am supposed to draw does not yet exist, since only fragments of it are visible, some parts of this narrative may be inconsistent with others. Like Cervantes' tale of Sancho Panza, who loses his donkey in one chapter but a few chapters later, thanks to the forgetfulness of the author, is riding the dear little animal again, our story has contradictions. We must start with the well-established concept that matter is composed of corpuscles or atoms, whose existence has been quite "tangibly" demonstrated by many beautiful experiments, and with Max Planck's discovery that energy also comes in indivisible units, called quanta, which are supposed to be transferred abruptly from one carrier to another.

But then Sancho Panza's donkey will return. For I shall have to ask you to believe neither in corpuscles as permanent individuals nor in the suddenness of the transfer of an energy quantum. Discreteness is present, but not in the traditional sense of discrete single particles, let alone in the sense of abrupt processes. Discreteness arises merely as a structure from the laws governing the phenomena. These laws are by no means fully understood; a probably correct analogue from the physics of palpable bodies is the way various partial tones of a bell derive from its shape and from the laws of elasticity to which, of themselves, nothing discontinuous adheres.

EDITOR'S NOTE

This article is condensed from a lecture entitled "Our Conception of Matter," given by Professor Schrödinger in 1952 at a conference in Geneva organized by Rencontres Internationales de Genève. The condensation is based on a translation by Sonja Bargmann, and it is published here with the kind permission of Editions de la Baconnière of Neuchâtel, Switzerland, who are publishing the full lecture in a volume called *L'homme devant la science*, presenting the proceedings of the conference.

The idea that matter is made up of ultimate particles was advanced as early as the fifth century B.C. by Leucippus and Democritus, who called these particles atoms. The corpuscular theory of matter was lifted to physical reality in the theory of gases developed during the 19th century by James Clerk Maxwell and Ludwig Boltzmann. The concept of atoms and molecules in violent motion, colliding and rebounding again and again, led to full comprehension of all the properties of gases: their elastic and thermal properties, their viscosity, heat conductivity and diffusion. At the same time it led to a firm foundation of the mechanical theory of heat, namely, that heat is the motion of these ultimate particles, which becomes increasingly violent with rising temperature.

Within one tremendously fertile decade at the turn of the century came the discoveries of X-rays, of electrons, of the emission of streams of particles and other forms of energy from the atomic nucleus by radioactive decay, of the electric charges on the various particles. The masses of these particles, and of the atoms themselves, were later measured very precisely, and from this was discovered the mass defect of the atomic nucleus as a whole. The mass of a nucleus is less than the sum of the masses of its component particles; the lost mass becomes the binding energy holding the nucleus firmly together. This is called the packing effect. The nuclear forces of

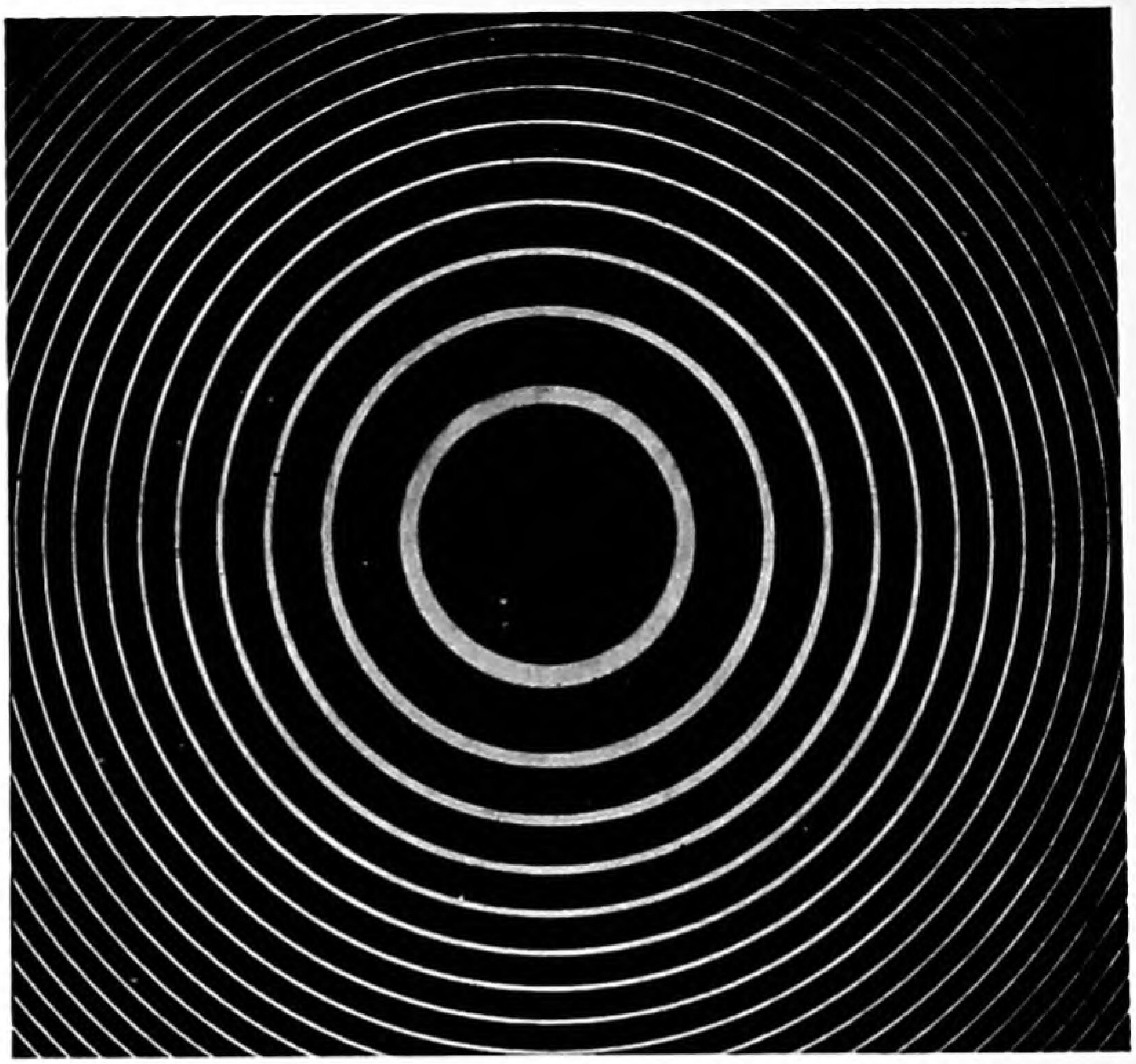
course are not electrical forces—those are repellent—but are much stronger and act only within very short distances, about 10^{-13} centimeter.

Here I am already caught in a contradiction. Didn't I say at the beginning that we no longer assume the existence of force fields apart from matter? I could easily talk myself out of it by saying: Well, the force field of a particle is simply considered a part of it. But that is not the fact. The established view today is rather that everything is at the same time both particle and field. Everything has the continuous structure with which we are familiar in fields, as well as the discrete structure with which we are equally familiar in particles. This concept is supported by innumerable experimental facts and is accepted in general, though opinions differ on details, as we shall see.

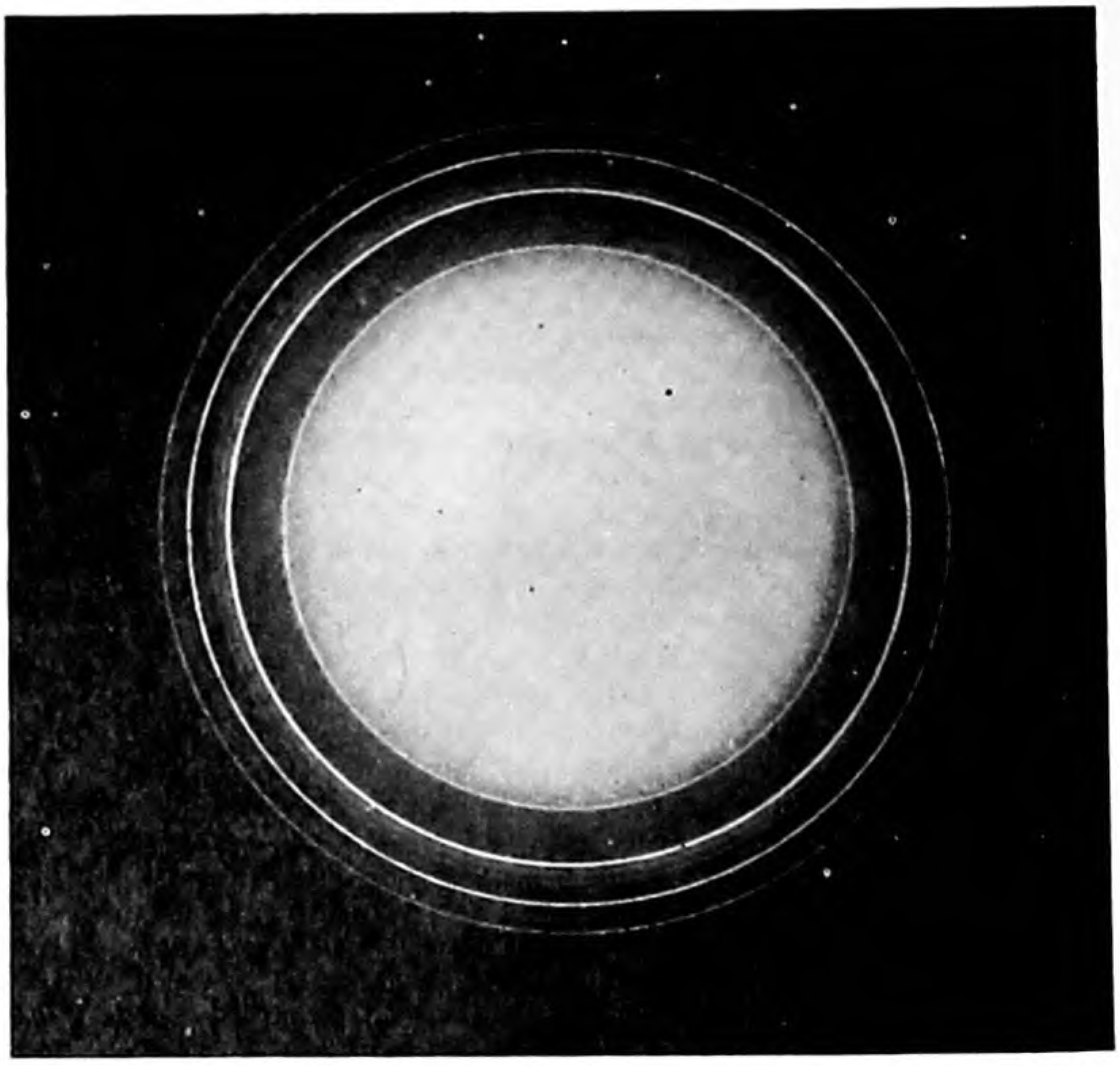
In the particular case of the field of nuclear forces, the particle structure is more or less known. Most likely the continuous force field is represented by the so-called pi mesons. On the other hand, the protons and neutrons, which we think of as discrete particles, indisputably also have a continuous wave structure, as is shown by the interference patterns they form when diffracted by a crystal. The difficulty of combining these two so very different character traits in one mental picture is the main stumbling-block that causes our conception of matter to be so uncertain.

Neither the particle concept nor the wave concept is hypothetical. The tracks in a photographic emulsion or in a Wilson cloud chamber leave no doubt of the behavior of particles as discrete units. The artificial production of nuclear particles is being attempted right now with terrific expenditure, defrayed in the main by the various state ministries of defense. It is true that one cannot kill anybody with one such racing particle, or else we should all be dead by now. But their study promises, indirectly, a hastened realization of the plan for the annihilation of mankind which is so close to all our hearts.

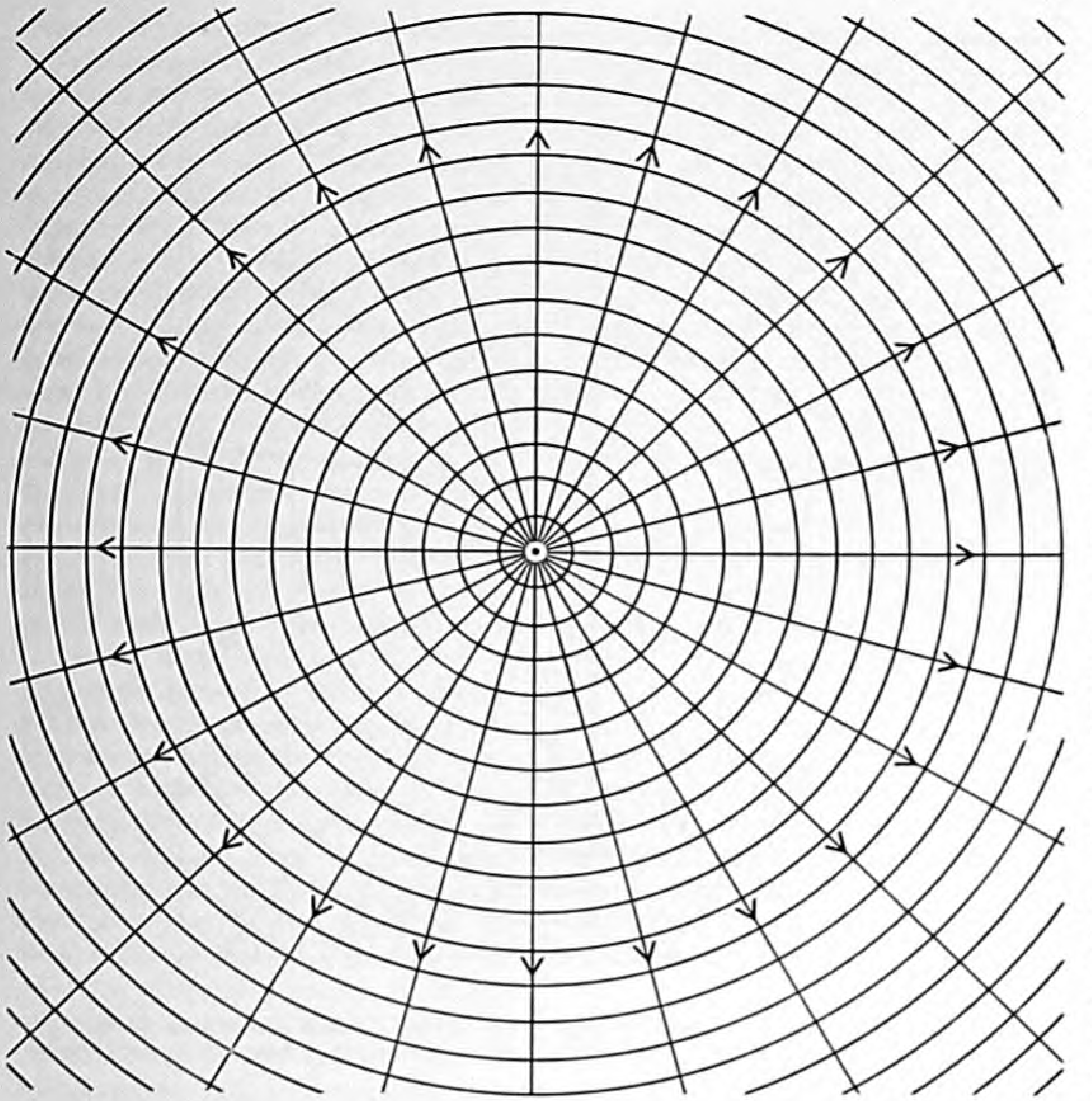
You can easily observe particles yourself by looking at a luminous numeral of your wrist watch in the dark with a magnifying glass. The luminosity surges and undulates, just as a lake sometimes twinkles in the sun. The light consists of sparklets, each produced by a so-called alpha particle (helium nucleus) expelled by a radioactive atom which in this process is transformed into a different atom. A specific device for detecting and recording single particles is the



LIGHT INTERFERENCE pattern, showing the wave nature of light, was produced at the National Bureau of Standards, using light from mercury vapor and an interferometer.



ELECTRON INTERFERENCE pattern from a crystal diffraction experiment at the Radio Corporation of America Laboratories gives convincing evidence that electrons are waves.



WAVE DIAGRAM in two dimensions shows wave fronts (*circles*) and wave "normals" or "rays" (*arrows*). In three dimensions the fronts would be surfaces like layers in an onion.

Geiger-Müller counter. In this short résumé I cannot possibly exhaust the many ways in which we can observe single particles.

Now to the continuous field or wave character of matter. Wave structure is studied mainly by means of diffraction and interference—phenomena which occur when wave trains cross each other. For the analysis and measurement of light waves the principal device is the ruled grating, which consists of a great many fine, parallel, equidistant lines, closely engraved on a specular metallic surface. Light impinging from one direction is scattered by them and collected in different directions depending on its wavelength. But even the finest ruled gratings we can produce are too coarse to scatter the very much shorter waves associated with matter. The fine lattices of crystals, however, which Max von Laue first used as gratings to analyze the very short X-rays, will do the same for "matter waves." Directed at the surface of a crystal, high-velocity streams of particles manifest their wave nature. With crystal gratings physicists have diffracted and measured the wavelengths of electrons, neutrons and protons.

What does Planck's quantum theory have to do with all this? Planck told us in 1900 that he could comprehend the radiation from red-hot iron, or from an incandescent star such as the sun, only if this radiation was produced in discrete portions and transferred in such discrete quantities from one carrier to another (*e.g.*, from atom to atom). This was extremely startling, because up to that time energy had been a highly abstract concept. Five years later Einstein told us that energy has mass and mass is energy; in other words, that they are one and the same. Now the scales begin to fall from our eyes: our dear old atoms, corpuscles, particles are Planck's energy quanta. *The carriers of those quanta are themselves quanta.* One gets dizzy. Something quite fundamental must lie at the bottom of this, but it is not surprising that the secret is not yet understood. After all, the scales did not fall suddenly. It took 20 or 30 years. And perhaps they still have not fallen completely.

The next step was not quite so far-reaching, but important enough. By an ingenious and appropriate generalization of Planck's hypothesis Niels Bohr taught us to understand the line spectra of atoms and molecules and how atoms were composed of heavy, positively charged nuclei with light, negatively charged electrons revolving around them. Each small system—atom or molecule—can harbor only definite discrete energy quantities, corresponding to its nature or its constitution. In transition from a higher to a lower "energy level" it emits the excess energy as a radiation quantum of definite wavelength, inversely proportional to the quantum given off. This means that a quantum of given magnitude manifests itself in a periodie process of definite frequency which is directly proportional to the quantum; the frequency equals the energy quantum divided by the famous Planck's constant, h.

According to Einstein a particle has the energy mc^2, m being the mass of the particle and c the velocity of light. In 1925 Louis de Broglie drew the inference, which rather suggests itself, that a particle might have associated with it a wave process of frequency mc^2 divided by h. The particle for which he postulated such a wave was the electron. Within two years the "electron waves" required by his theory were demonstrated by the famous electron diffraction experiment of C. J. Davisson and L. H. Germer. This was the starting point for the cognition that everything—anything at all—is simultaneously particle and wave field. Thus de Broglie's dissertation initiated our uncertainty about the nature of matter. Both the particle picture and the wave picture have truth value, and we cannot give up either one or the other. But we do not know how to combine them.

That the two pictures are connected is known in full generality with great precision and down to amazing details. But concerning the unification to a single, concrete, palpable picture opinions are so strongly divided that a great many deem it altogether impossible. I shall briefly sketch the connection. But do not expect that a uniform, concrete picture will emerge before you; and do not blame the lack of success either on my ineptness in exposition or your own denseness—nobody has yet succeeded.

One distinguishes two things in a wave. First of all, a wave has a front,

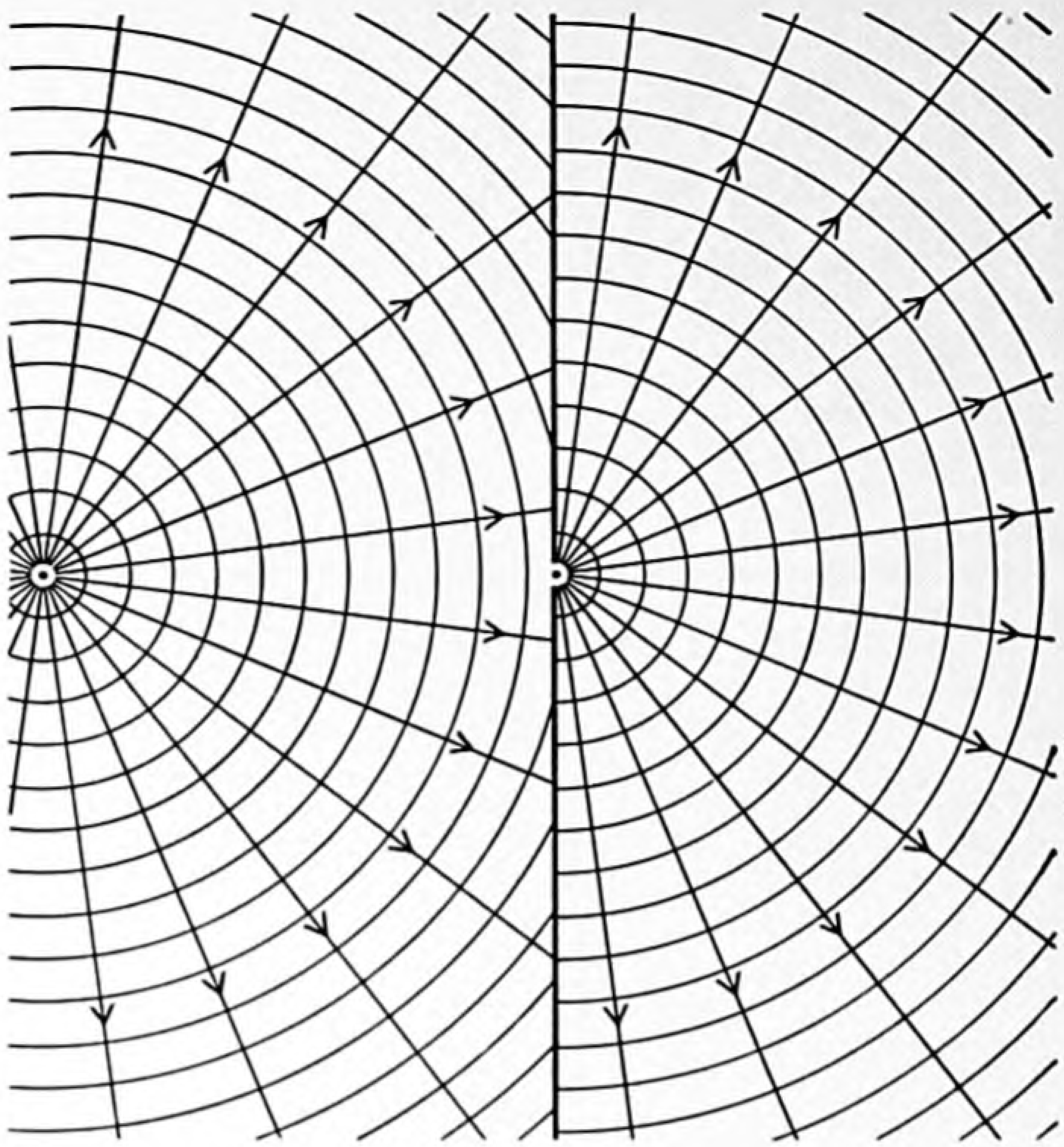
and a succession of wave fronts forms a system of surfaces like the layers of an onion. You are familiar with the two-dimensional analogue of the beautiful wave circles that form on the smooth surface of a pond when a stone is thrown in. The second characteristic of a wave, less intuitive, is the path along which it travels—a system of imagined lines perpendicular to the wave fronts. These lines are known as the wave "normals" or "rays."

We can make the provisional assertion that these rays correspond to the trajectories of particles. Indeed, if you cut a small piece out of a wave, approximately 10 or 20 wavelengths along the direction of propagation and about as much across, such a "wave packet" would actually move along a ray with exactly the same velocity and change of velocity as we might expect from a particle of this particular kind at this particular place, taking into account any force fields acting on the particle.

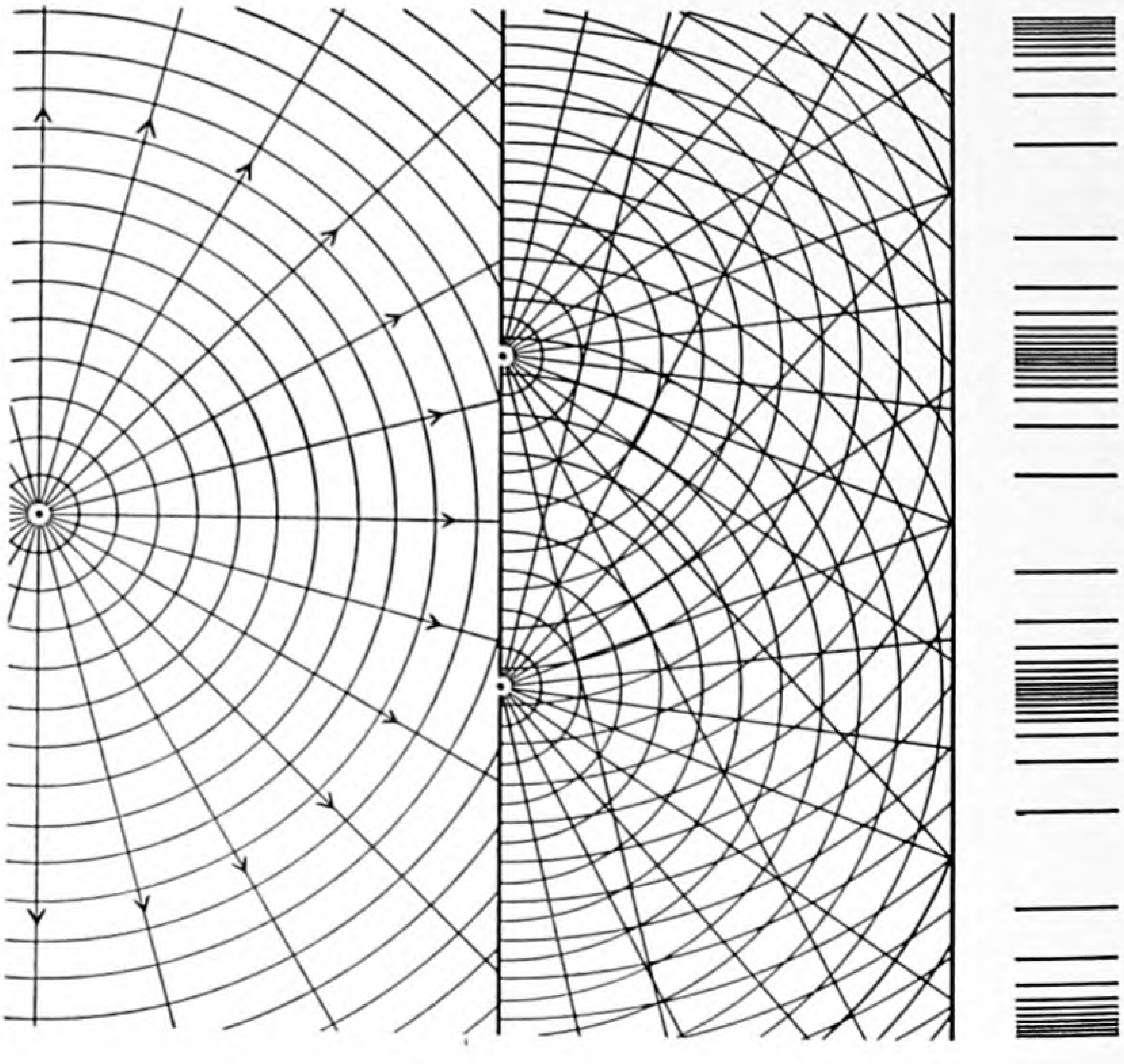
Here I falter. For what I must say now, though correct, almost contradicts this provisional assertion. Although the behavior of the wave packet gives us a more or less intuitive picture of a particle, which can be worked out in detail (*e.g.*, the momentum of a particle increases as the wavelength decreases; the two are inversely proportional), yet for many reasons we cannot take this intuitive picture quite seriously. For one thing, it is, after all, somewhat vague, the more so the greater the wavelength. For another, quife often we are dealing not with a small packet but with an extended wave. For still another, we must also deal with the important special case of very small "packelets" which form a kind of "standing wave" which can have no wave fronts or wave normals.

One interpretation of wave phenomena which is extensively supported by experiments is this: At each position of a uniformly propagating wave train there is a twofold structural connection of interactions, which may be distinguished as "longitudinal" and "transversal." The transversal structure is that of the wave fronts and manifests itself in diffraction and interference experiments; the longitudinal structure is that of the wave normals and manifests itself in the observation of single particles. However, these concepts of longitudinal and transversal structures are not sharply defined and absolute, since the concepts of wave front and wave normal are not, either.

The interpretation breaks down completely in the special case of the standing



DIFFRACTION is characteristic of waves. When a wave (*left*) comes to a barrier perforated with a small hole, it diffracts around the edges of the hole to form a new wave (*right*).



INTERFERENCE is also evidence of waves. Its characteristic pattern is formed when rays interact. For light waves the pattern shows up as bright and dark bands on a screen (*right*).

waves mentioned above. Here the whole wave phenomenon is reduced to a small region of the dimensions of a single or very few wavelengths. You can produce standing water waves of a similar nature in a small basin if you dabble with your finger rather uniformly in its center, or else just give it a little push so that the water surface undulates. In this situation we are not dealing with uniform wave propagation; what catches the interest are the normal frequencies of these standing waves. The water waves in the basin are an analogue of a wave phenomenon associated with electrons, which occurs in a region just about the size of the atom. The normal frequencies of the wave group washing around the atomic nucleus are universally found to be exactly equal to Bohr's atomic "energy levels" divided by Planck's constant *h*. Thus the ingenious yet somewhat artificial assumptions of Bohr's model of the atom, as well as of the older quantum theory in general, are superseded by the far more natural idea of de Broglie's wave phenomenon. The wave phenomenon forms the "body" proper of the atom. It takes the place of the individual pointlike electrons which in Bohr's model are supposed to swarm around the nucleus. Such pointlike single particles are completely out of the question within the atom, and if one still thinks of the nucleus itself in this way one does so quite consciously for reasons of expediency.

What seems to me particularly important about the discovery that "energy levels" are virtually nothing but the frequencies of normal modes of vibration is that now one can do without the assumption of sudden transitions, or quantum jumps, since two or more normal modes may very well be excited simultaneously. The discreteness of the normal frequencies fully suffices—so I believe—to support the considerations from which Planck started and many similar and just as important ones—I mean, in short, to support all of quantum thermodynamics.

The theory of quantum jumps is becoming more and more inacceptable, at least to me personally, as the years go on. Its abandonment has, however, far-reaching consequences. It means that one must give up entirely the idea of the exchange of energy in well-defined quanta and replace it with the concept of resonance between vibrational frequencies. Yet we have seen that because of the identity of mass and energy, we must consider the particles themselves as Planck's energy quanta. This is at first frightening. For the substituted theory implies that we can no longer consider the individual particle as a well-defined permanent entity.

That it is, in fact, no such thing can be reasoned in other ways. For one thing, there is Werner Heisenberg's famous uncertainty principle, according to which a particle cannot simultaneously have a well-defined position and a sharply defined velocity. This uncertainty implies that we cannot be sure that the same particle could ever be observed twice. Another conclusive reason for not attributing identifiable sameness to individual particles is that we must obliterate their individualities whenever we consider two or more interacting particles of the same kind, *e.g.*, the two electrons of a helium atom. Two situations which are distinguished only by the interchange of the two electrons must be counted as one and the same; if they are counted as *two* equal situations, nonsense obtains. This circumstance holds for any kind of particle in arbitrary numbers without exception.

Most theoreticians will probably accept the foregoing reasoning and admit that the individual particle is not a well-defined permanent entity of detectable identity or sameness. Nevertheless this inadmissible concept of the individual particle continues to play a large role in their ideas and discussions. Even deeper rooted is the belief in "quantum jumps," which is now surrounded with a highly abstruse terminology whose common-sense meaning is often difficult to grasp. For instance, an important word in the standing vocabulary of quantum theory is "probability," referring to transition from one level to another. But, after all, one can speak of the probability of an event only assuming that, occasionally, it actually occurs. If it does occur, the transition must indeed be sudden, since intermediate stages are disclaimed. Moreover, if it takes time, it might conceivably be interrupted halfway by an unforeseen disturbance. This possibility leaves one completely at sea.

The wave *v.* corpuscle dilemma is supposed to be resolved by asserting that the wave field merely serves for the computation of the probability of finding a particle of given properties at a given position if one looks for it there. But once one deprives the waves of reality and assigns them only a kind of informative role, it becomes very difficult to understand the phenomena of interference and diffraction on the basis of the combined action of discrete single particles. It certainly seems easier to explain particle tracks in terms of waves than to explain the wave phenomenon in terms of corpuscles.

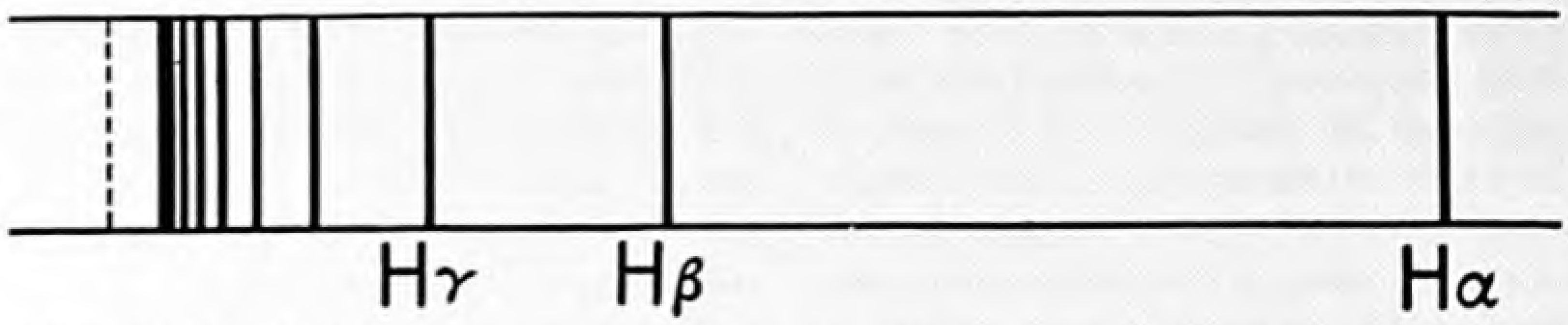
"Real existence" is, to be sure, an expression which has been virtually chased to death by many philosophical hounds. Its simple, naive meaning has almost become lost to us. Therefore I want to recall something else. I spoke of a corpuscle's not being an individual. Properly speaking, one never observes the same particle a second time—very much as Heraclitus says of the river. You cannot mark an electron, you cannot paint it red. Indeed, you must not even *think* of it as marked; if you do, your "counting" will be false and you will get wrong results at every step—for the structure of line spectra, in thermodynamics and elsewhere. A wave, on the other hand, can easily be imprinted with an individual structure by which it can be recognized beyond doubt. Think of the beacon fires that guide ships at sea. The light shines according to a definite code; for example: three seconds light, five seconds dark, one second light, another pause of five seconds, and again light for three seconds—the skipper knows that is San Sebastian. Or you talk by wireless telephone with a friend across the Atlantic; as soon as he says, "Hello there, Edward Meier speaking," you know that his voice has imprinted on the radio wave a structure which can be distinguished from any other. But one does not have to go that far. If your wife calls, "Francis!" from the garden, it is exactly the same thing, except that the structure is printed on sound waves and the trip is shorter (though it takes somewhat longer than the journey of radio waves across the Atlantic). All our verbal communication is based on imprinted individual wave structures. And, according to the same principle, what a wealth of details is transmitted to us in rapid succession by the movie or the television picture!

This characteristic, the individuality of the wave phenomenon, has already been found to a remarkable extent in the very much finer waves of particles. One example must suffice. A limited volume of gas, say helium, can be thought of either as a collection of many helium atoms or as a superposition of elementary wave trains of matter waves. Both views lead to the same theoretical results as to the behavior of the gas upon

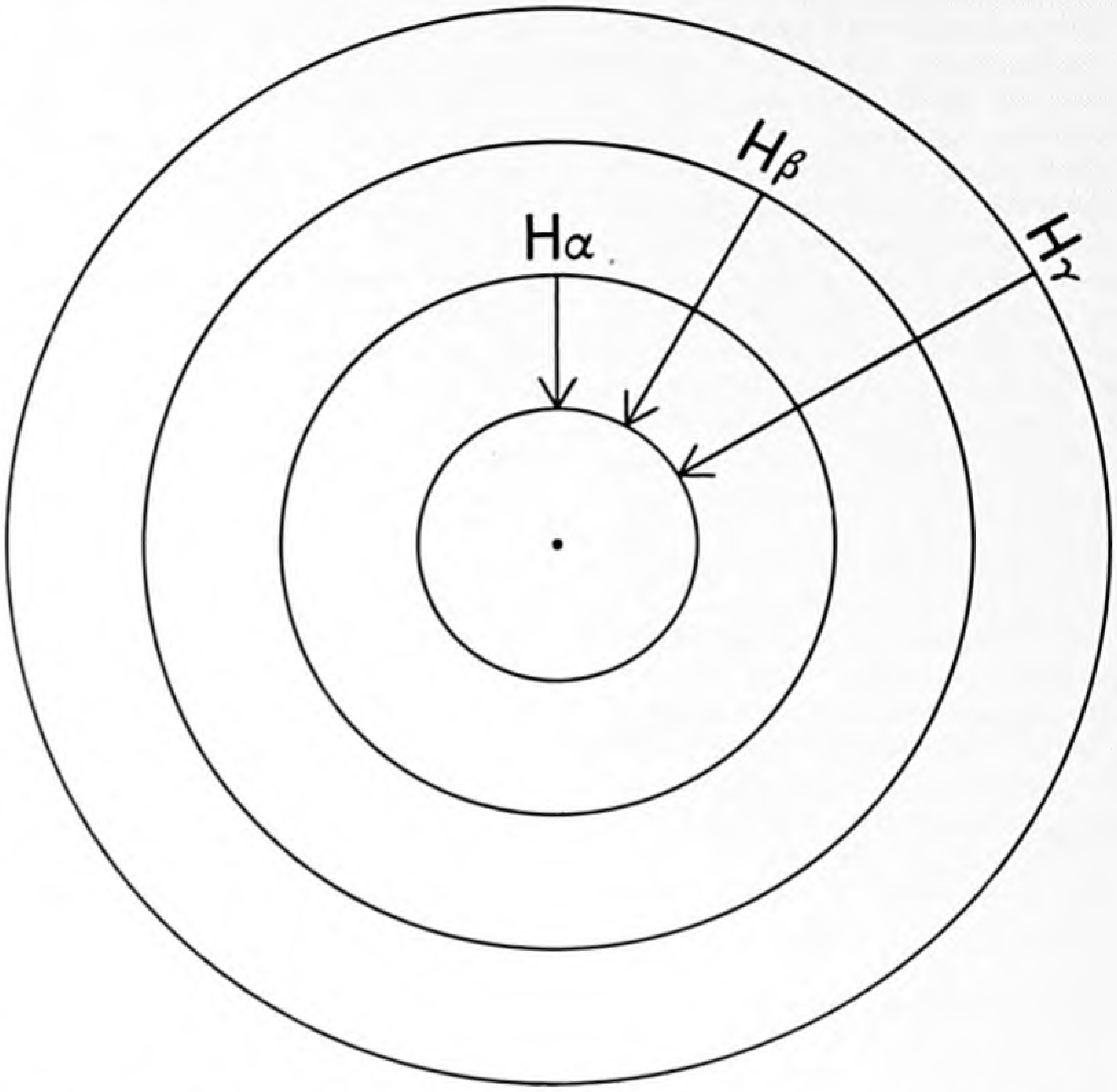
heating, compression, and so on. But when you attempt to apply certain somewhat involved enumerations to the gas, you must carry them out in different ways according to the mental picture with which you approach it. If you treat the gas as consisting of particles, then no individuality must be ascribed to them, as I said. If, however, you concentrate on the matter wave trains instead of on the particles, every one of the wave trains has a well-defined structure which is different from that of any other. It is true that there are many pairs of waves which are so similar to each other that they could change roles without any noticeable effect on the gas. But if you should count the very many similar states formed in this way as merely a single one, the result would be quite wrong.

In spite of everything we cannot completely banish the concepts of quantum jump and individual corpuscle from the vocabulary of physics. We still require them to describe many details of the structure of matter. How can one ever determine the weight of a carbon nucleus and of a hydrogen nucleus, each to the precision of several decimals, and detect that the former is somewhat lighter than the 12 hydrogen nuclei combined in it, without accepting for the time being the view that these particles are something quite concrete and real? This view is so much more convenient than the roundabout consideration of wave trains that we cannot do without it, just as the chemist does not discard his valence-bond formulas, although he fully realizes that they represent a drastic simplification of a rather involved wave-mechanical situation.

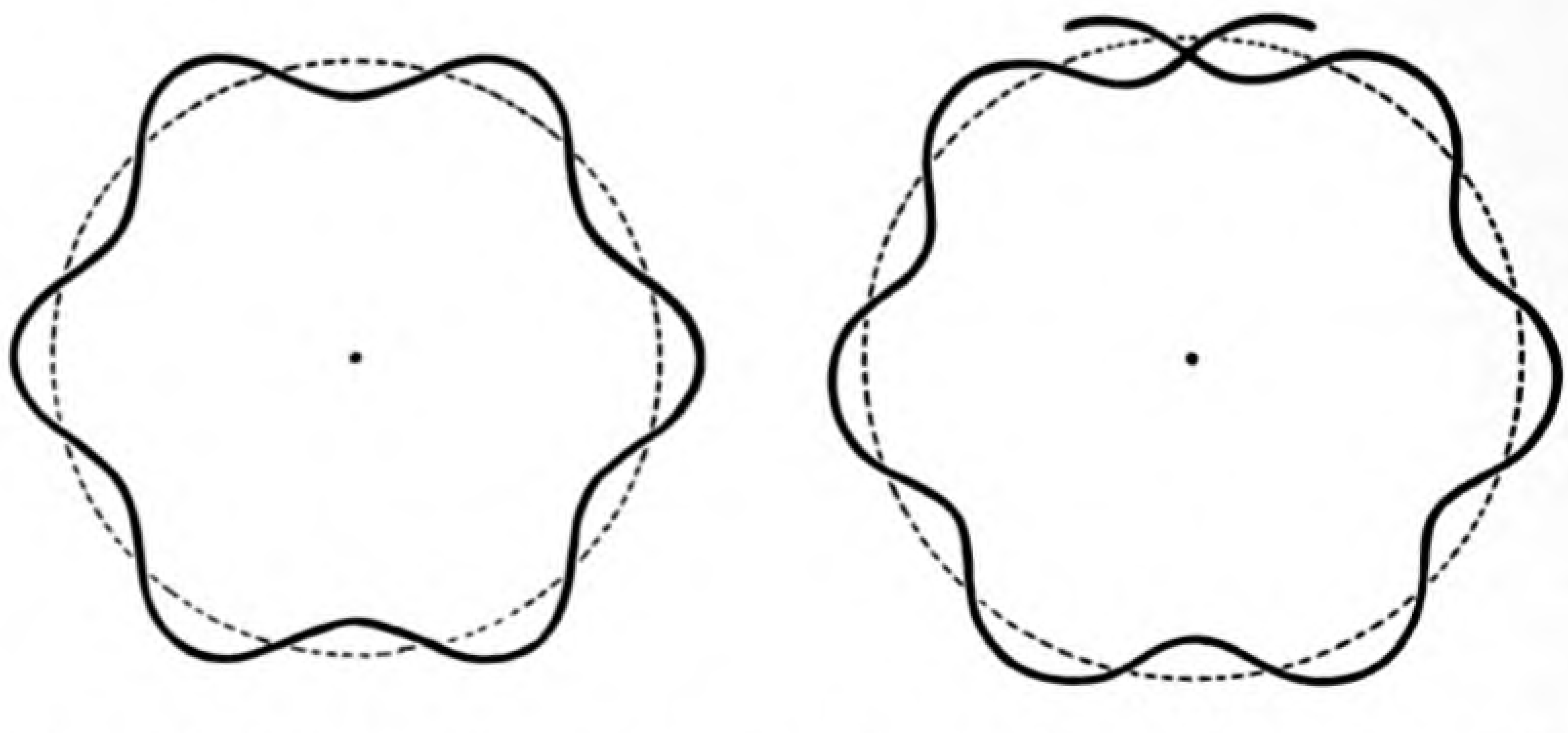
If you finally ask me: "Well, what *are* these corpuscles, really?" I ought to confess honestly that I am almost as little prepared to answer that as to tell where Sancho Panza's second donkey came from. At the most, it may be permissible to say that one can think of particles as more or less temporary entities within the wave field whose form and general behavior are nevertheless so clearly and sharply determined by the laws of waves that many processes take place *as if* these temporary entities were substantial permanent beings. The mass and the charge of particles, defined with such precision, must then be counted among the structural elements determined by the wave laws. The conservation of charge and mass in the large must be considered as a statistical effect, based on the "law of large numbers."



HYDROGEN SPECTRUM expresses the behavior of a fundamental constituent of matter, the electron. Shown above is a part of the Balmer series of spectral lines, which are in the visible light range. Each line is the result of a change in energy of the atom's electron.



BOHR THEORY explained spectral lines of hydrogen by postulating a pointlike electron revolving around the nucleus in any of a number of possible orbits. In falling from one to another, the electron emits light energy whose wavelength is that of one of the spectral lines.



WAVE MECHANICS sees the electron not as a point mass, but as a standing wave washing to and fro in the atom. Some modes of vibration are possible (*left*), while others are not (*right*). The possible modes correspond exactly to the Bohr theory's possible energy levels.

The Author

ERWIN SCHRÖDINGER is one of the founders of modern physics. For developing the theory of wave mechanics he shared the Nobel prize in 1933 with the British physicist P. A. M. Dirac. Schrödinger, born in Vienna in 1887, comes from the distinguished Austrian school of physics which produced Ernst Mach and Ludwig Boltzmann. He succeeded Max Planck in the chair of theoretical physics at the University of Berlin in 1927. Upon Hitler's rise to power he went to Dublin to join the Institute for Advanced Study, where he is now. In recent years he has been trying to combine the field theories of physics into a unified structure. He has also been interested in more general unifications of science, and perhaps his most famous book is *What Is Life?*

Bibliography

SCIENCE AND HUMANISM: PHYSICS IN OUR TIME. Erwin Schrödinger. Cambridge University Press, 1951.

THE REVOLUTION IN PHYSICS. Louis de Broglie. The Noonday Press, 1953.

SCIENTIFIC AMERICAN April 1950, Vol. 182, No. 4, pp. 38-47 OFFPRINT 242

THE SYNTHETIC ELEMENTS I

by Glenn T. Seaborg and I. Perlman

Wartime studies in the chemistry of fissionable materials led to a wholesale synthesis of unknown elements. Four of them filled gaps in the periodic table of 92 elements; five extended it beyond uranium.

THE urge to take our material world apart and identify its ultimate units of construction is at least as old as the early Greek philosophers. We have come a long way from their conclusion that all substances are variations of an Olympian brew composed of only four ingredients—fire, water, earth and air—but the end of the quest is not yet in sight. At the moment the list of identified elements stands at 97. The physicists, of course, have broken these down into protons, neutrons and units which may be even more basic. In this article, however, we shall stop short of the subatomic world and confine ourselves to the elements, considering protons and neutrons only as they affect the elements' stability. From the strictly chemical point of view, the elements may be considered the blocks of which our universe is built.

Our particular concern in this article will be the so-called synthetic elements, which for all practical purposes are not found in nature but are created only by the alchemy of the modern laboratory. Let us make clear at the beginning what we mean when we say that the synthetic elements are missing in nature. Actually tiny amounts of some of them, such as plutonium, have been detected in the earth, and all of them doubtless existed in considerable amounts at the primordial creation of the elements. But without exception they are so unstable that the original atoms must have disappeared long ago; any such atoms now found in nature are created only rarely by spontaneous nuclear reactions due to cosmic-ray bombardment or natural radioactivity.

The idea that all matter could be reduced to a limited number of chemically indivisible elements began to take form in the 19th century. It developed principally from the discovery of certain periodic similarities among the known elements. When the elements were arranged in the sequence of their atomic weights, it was found that elements occurring at certain intervals on the list resembled one another in chemical properties. This resulted in the construction of a periodic table of the elements, which in turn disclosed some gaps. It was logical to assume that the missing chemical properties should be attributed to still undetected elements.

With the discovery of X-rays and of the atomic nucleus at the turn of the century, a more meaningful picture of the differences among elements began to emerge. We now know that the distinguishing mark of each element—what determines its chemical properties—is the number of electrons it possesses, and that this in turn is uniquely determined by the number of positive charges or protons in the nucleus. The electrons are attached to the nucleus in successive shells, and as we go up the periodic table from the lower to the higher elements we find that the electrons closest to the nucleus are attached more and more firmly to the atom. The dislodging of one of these inner electrons is immediately followed by an outer electron falling into the vacancy, and the energy released by this event appears as an X-ray. The wavelength of the X-ray is characteristic of the element. It was H. G. J. Moseley of England who discovered this relationship and thereby was able to arrange the elements according to atomic number and to tell precisely which elements were missing. Gradually most of the gaps between hydrogen (atomic number 1) and uranium (atomic number 92) were filled by the discovery of new elements. By 1925 only four elements remained to be found: those of atomic numbers 43, 61, 85 and 87.

As one might expect, a new element did not necessarily appear for the first time in pure form, nor did it assert its singularity. Some research workers may have handled substances in relatively pure form and failed to recognize them as new elements. More often new elements were reported which proved to be identical with previously known elements or mixtures. During the 1920s and 1930s a number of workers reported the discovery of elements 43, 61, 85 and 87. They gave these elements such names as masurium, illinium, florentium, alabamine, virginium and moldavium, and to this day these names appear in some tables of elements. It is fairly certain, however, that none of these elements can exist in nature in quantities detectable by the methods of investigation then employed. Actually the unambiguous identification of three of the four elements (the exception: element 87) had to await their preparation by artificial means, and even element 87 can be prepared more readily by transmutation than from natural sources.

Stable and Unstable Isotopes

To understand the transmutation of elements, we must turn to considerations of nuclear stability. An element, as we have noted, is uniquely characterized by the number of protons in the nucleus. But the number of neutrons associated with a given number of protons may vary. This results in the existence of various species of the same element, known as isotopes, which differ in weight and stability but not appreciably in chemical properties. Relatively few of the possible isotopes of any element are stable; in fact, of the 1,000 isotopes of the 97 elements known to date, only about 275 are stable. Several hundred unstable nuclei

may yet be prepared, but probably few stable nuclei remain to be discovered. The paucity of stable nuclei is largely explained by the interconversion of protons and neutrons; a proton will change into a neutron or *vice versa* if there is a slight imbalance from a norm characteristic of each region of the periodic table. The nucleus is much more stable if it has an even number of neutrons or protons. As a result each element with an odd number of protons has only one or two stable isotopes.

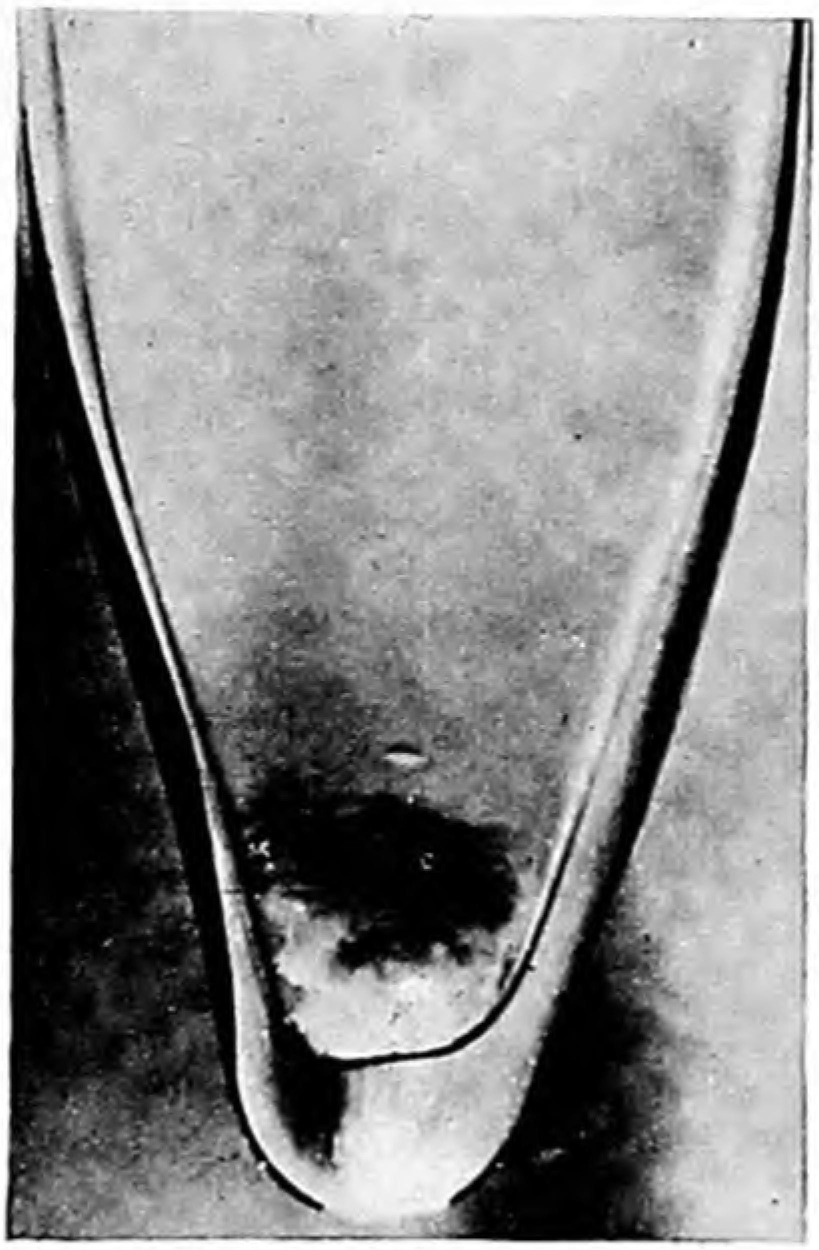
The margin by which an isotope may be stable or unstable is indeed small when compared with the total amount of energy involved in binding together the components of a nucleus. In moderately heavy nuclei the total binding energy is about 1,000 million electron volts, yet if one nucleus is bound more firmly by even .01 mev than another with one more proton and one less neutron, the second nucleus will decay into the first. Such small irregularities in nuclear binding may deprive some odd-numbered elements of the possibility of having even one stable isotope. Consequently such an element, barring some freakish factor that prevented its most nearly stable isotope from decaying, would not be found in nature.

It may be of interest to speculate how different our lives would be, if indeed we would be here at all, should certain elements be unstable. One element with only a single stable isotope, for example, is iodine. This element, as a constituent of thyroxine, the hormone of the thyroid gland, is vital as a regulator of growth and development and of the metabolic rate. It is difficult to visualize what form vertebrate animal life would have taken had this element been missing. Another element with but one stable isotope is gold. As well as we can measure it, this isotope, Au^{197}, is considerably less than one mev more stable than the artificial and highly unstable mercury isotope Hg^{197}. If this situation were reversed, Fort Knox might still be used to store something, but it would not be gold.

The Missing Elements

Below lead, element 82, only two elements are missing in nature. These are elements 43 (technetium) and 61 (promethium) which, as we shall see, may be prepared artificially in radioactive form just as one may prepare radioactive isotopes of all of the elements. The form of instability responsible for the absence of elements 43 and 61 involves neutron-proton interconversions. It is called beta-instability, meaning that the nucleus emits beta particles, *i.e.*, electrons, in attaining stability.

Among the heavier elements another type of instability sets in. This type is a consequence of what may loosely be considered an overstuffing of positive

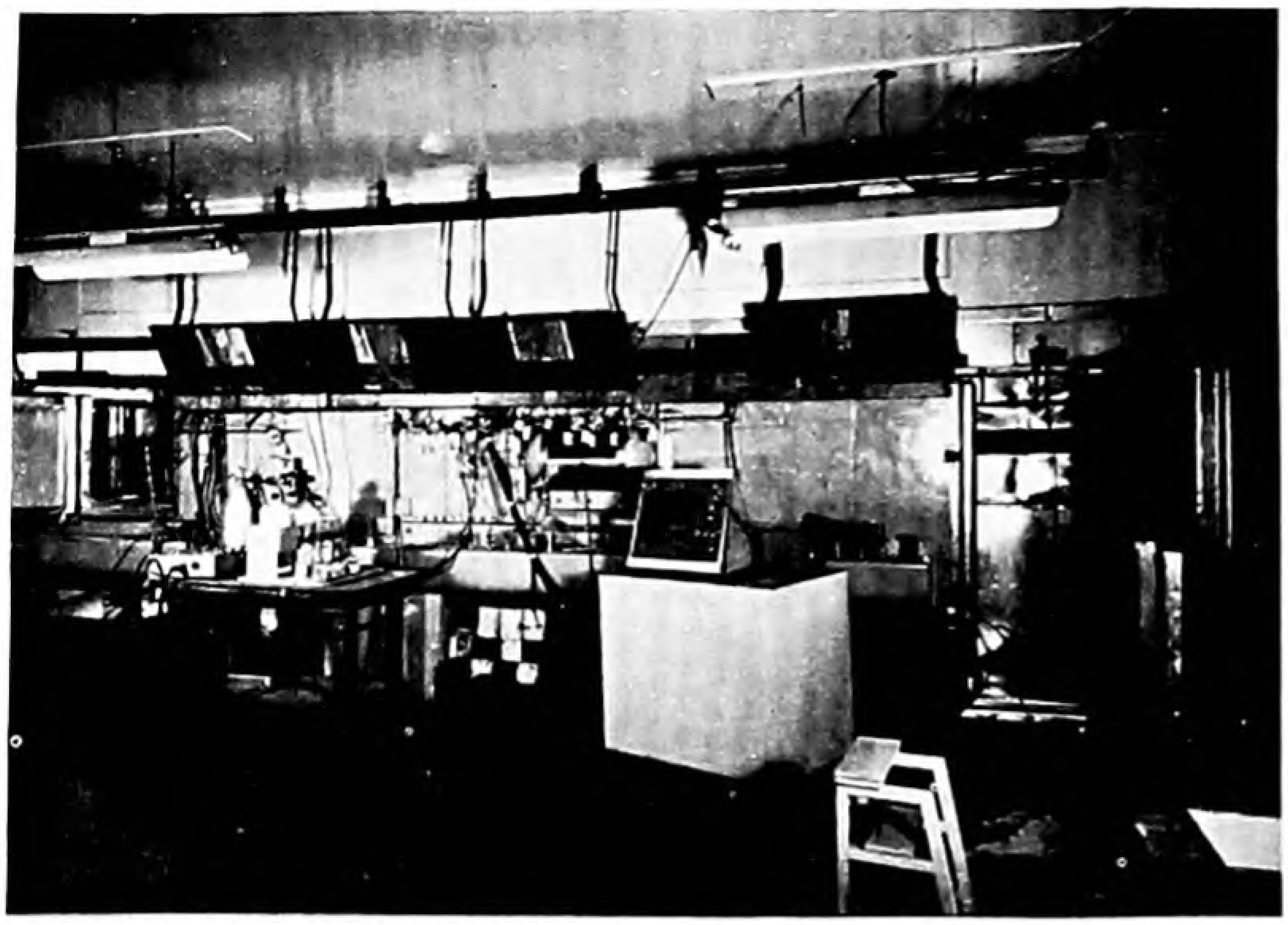


PLUTONIUM hydroxide is isolated in a few crystals at the bottom of a capillary tube. Sample was prepared at University of Chicago in 1942.

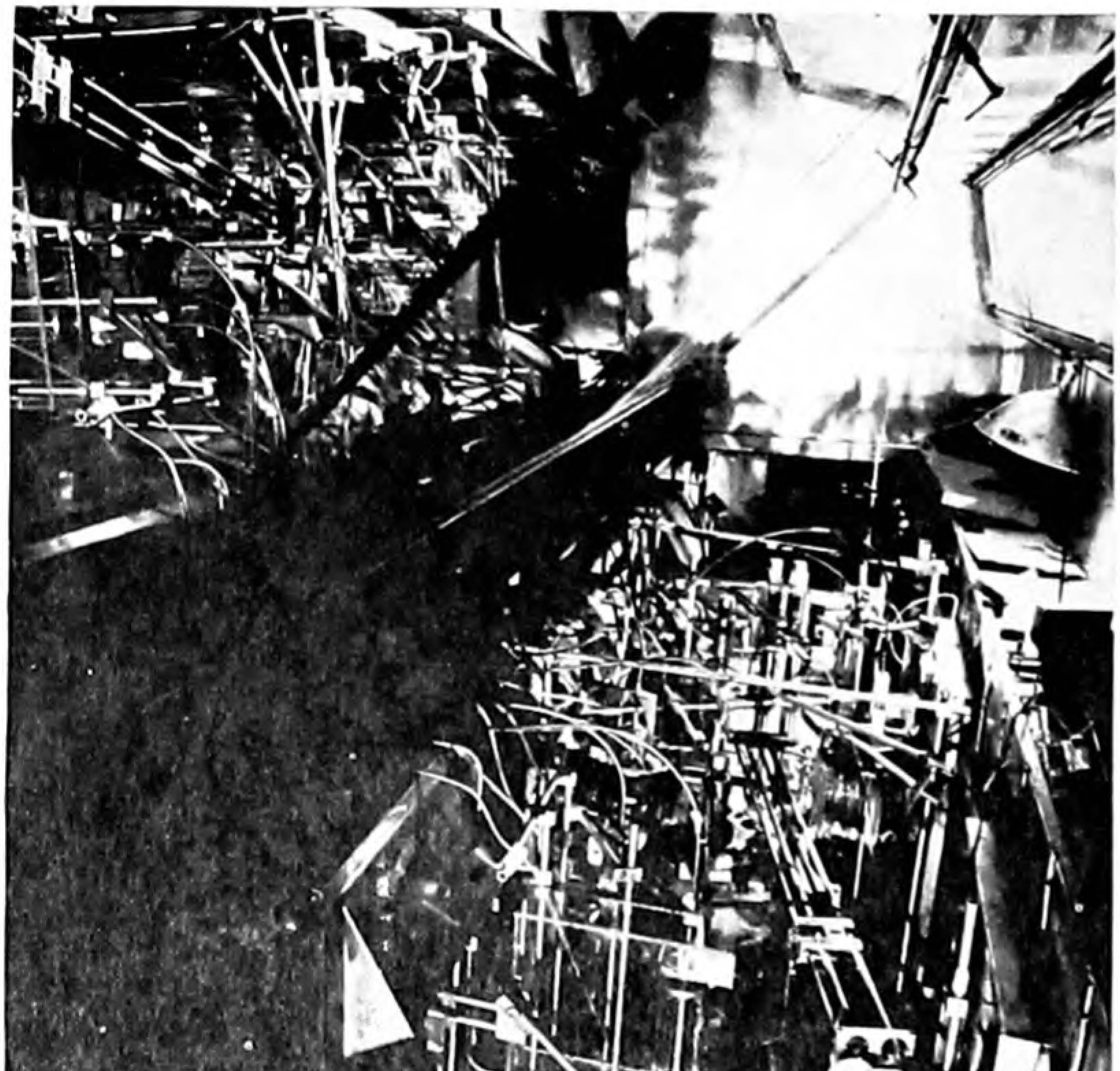
CURIUM is so intensely radioactive that it glows by its own light in a water solution. Curium was discovered at University of Chicago in 1944.

PROMETHIUM nitrate is isolated as clustered crystals. The crystals in the original photomicrograph have been enlarged by about 30 diameters.

TECHNETIUM is prepared in the compound ammonium pertechnetate. In original photomicrograph crystals were enlarged 16 diameters.



HOT LABORATORY at the University of California provides facilities for working with highly radioactive elements, synthetic and otherwise. Elements are chemically manipulated behind a lead wall by remote control (*below*).



BEHIND LEAD WALL is equipment for manipulating radioactive elements. At upper left is a mirror in which the manipulations may be observed. Some stages in the isolation of curium were carried out in this laboratory.

charge in the nucleus. The nucleus has an urge to get rid of protons. The mechanism for relieving its condition is the emission of alpha particles, or helium nuclei, composed of two protons and two neutrons. The reason this complex particle rather than a proton is emitted is simply that the helium nucleus is such a stable structure that it is more economical of energy to rid the nucleus of protons in this fashion than individually. By the time bismuth, element 83, is reached, alpha instability sets in as a general condition. (Some nuclei do not exhibit their alpha instability, however, because their beta-decay rate is much faster than their alpha-decay rate.) These alpha-emitters have vastly varying lifetimes: some as short as a microsecond, others comparable with the age of the earth. The point to be made here is that there are only three nuclei above bismuth sufficiently long-lived to have survived geological time—thorium 232, uranium 235 and uranium 238. These isotopes are responsible for the fact that the earth still has small amounts of the elements between 83 and 92, for the latter arise as products of the decay of uranium and thorium. Thus the existence of this small island of relative stability at thorium and uranium is the only factor preventing the termination of the periodic table at bismuth.

As these three nuclei decay, they maintain their various products in equilibrium with them, in amounts that depend on their relative stability or half-lives. For example, radium 226 is one of the decay products of U^{238}. Since the respective half-lives of these isotopes are 1,600 years and 4.5 billion years, the two are found together in the ratio of one part of radium to three million parts of uranium, or a third of a gram of radium to a ton of uranium. Two of the elements with extremely short half-lives, elements 85 and 87, are almost missed completely in the radioactive-decay series; element 87 occurs in uranium only in the fantastically low concentration of a few parts per billion billion. In the case of element 85, the amount that has been detected in nature is much smaller. Such small quantities, of course, cannot be isolated and are measurable only through their radioactivity.

It is only by the grace of an odd combination of unusual circumstances, by the way, that fissionable U^{235} still exists in the earth in sufficient quantity to provide us with nuclear chain reactors and atomic energy. U^{235} has a half-life of only .7 billion years, which means that more than 90 per cent has disappeared through radioactive decay during the three billion years of the earth's age. Thus only by the slenderest of margins does enough U^{235} remain in natural uranium to operate a nuclear reactor or to make its separation from U^{238} feasible. And this is only half the story. Recent studies of alpha radioactivity have

shown that U^{235} falls into a category of nuclei whose half-lives are longer than would be predicted from their decay energy—the principal factor influencing the half-life. Even among this group in which alpha decay is "forbidden," U^{235} is something of a freak. For its particular decay energy it might be expected to decay about 10 times more rapidly than it does; while even if its decay were only twice as rapid as it actually is, it would essentially have disappeared from the earth by this time.

Above uranium, alpha-decay half-lives again become quite short, so the transuranium elements of primordial origin are no longer present although there is every reason to believe that such elements were formed at the same time as the more stable ones. The longest-lived transuranium isotope known to date, neptunium 239, has a half-life of only two million years, which is almost 100 times too short to have permitted the element to persist through geological time.

Radiochemistry

What about the chemical behavior of these unstable elements? Basically the chemistry of a radioactive substance is no different from that it would have if it were not radioactive. There is one practical difference in handling it, however, and this is that we can detect it even when it is present in vanishingly small concentrations. If it were not for this facility, most unstable nuclei would remain undiscovered, for they are found or can be prepared only in unweighable amounts.

Today the techniques of the new branch of study called radiochemistry make it possible to obtain a great deal of information about chemical properties and to carry through chemical separations with amounts of material far too small to handle by the usual methods of chemistry. By the mere analysis of a substance's radioactivity it is possible to obtain semiquantitative information about its solubility, its oxidation-reduction potentials, its formation of complex ions and many other properties.

Once the element has been identified there is a great incentive for manufacturing it in visible amounts so that one can study its spectra, its crystal structure and many other properties that are inaccessible to radiochemical methods. With the advent of the nuclear reactor the synthesis of these radioactive elements is no longer difficult, provided the transmutations can be effected with neutrons. There are problems, however, in connection with the elements' great radioactivity. It is desirable to work with an isotope with a relatively long half-life, for if the half-life is short it may be difficult to produce the element at a faster rate than it decays. Also important is the hazard in handling these radioactive substances when they are made in visible amounts. A good example is the curium isotope Cm^{242}, the principal isotope of curium that has been used for experimentation up to this time. This isotope has a half-life of only five months, and if it were possible to make one milligram of it, its alpha radioactivity would be 3.5 curies. This would be a considerable amount of radioactivity to work with. If it were spread uniformly over the entire state of New York (we deliberately refrain from spreading it here in California), radioactivity could be detected on every square foot of ground.

Thus the experimenter has no alternative but to work with extremely small amounts of such isotopes. Ultramicrochemical methods have been developed, however, which permit almost any type of chemical and physical measurement to be made on only a few micrograms to a milligram of an element.

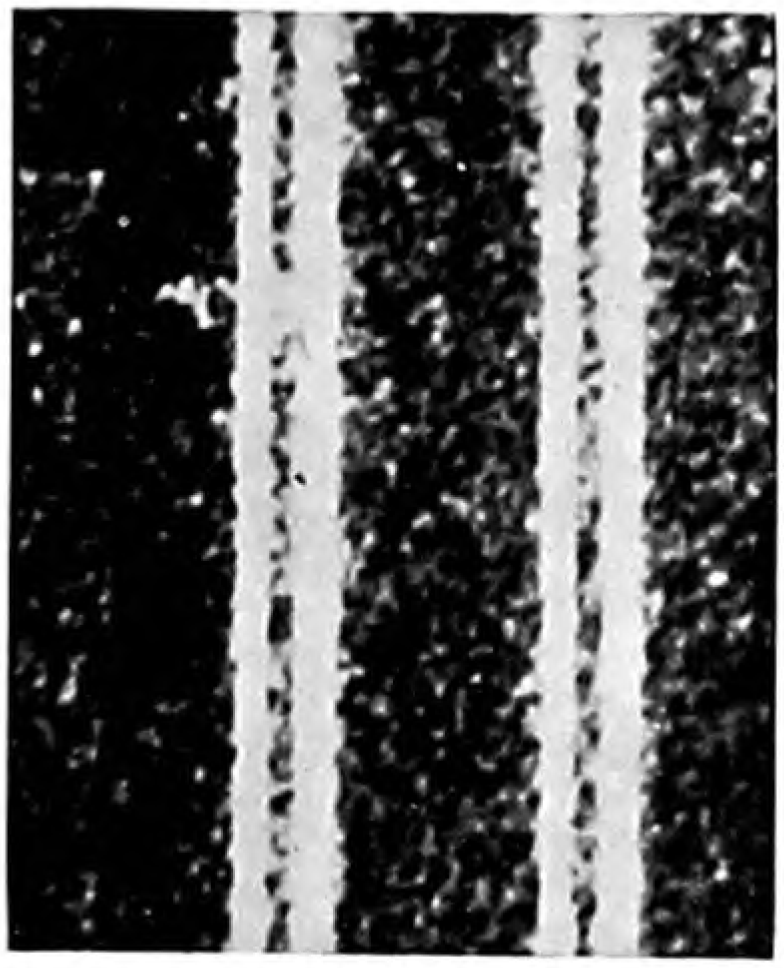
Element 43

The first synthetic element to be created was technetium, element 43, which filled the gap in the periodic table between molybdenum and ruthenium. Technetium was definitely identified for the first time by C. Perrier and E. Segrè of Italy in 1937. A sample of molybdenum that had been irradiated with deuterons in the University of California cyclotron was sent to them. From it they isolated a chemical fraction which, on the basis of its radioactive behavior, was distinct from all other known elements. Its chemistry conformed with what might have been expected from element 43, an element in the series known as Group VII. Such a group, as already indicated, is made up of elements in the periodic table that are chemically similar to one another because they have the same number of electrons in the outer shell. Further exploratory work showed that element 43 was somewhat closer in properties to rhenium, the next heaviest element in Group VII, than to manganese, the next lowest element in the group. Ten years later Segrè suggested that element 43 be named technetium (Tc), derived from the Greek *technikos*, signifying the element's artificial or "technical" origin.

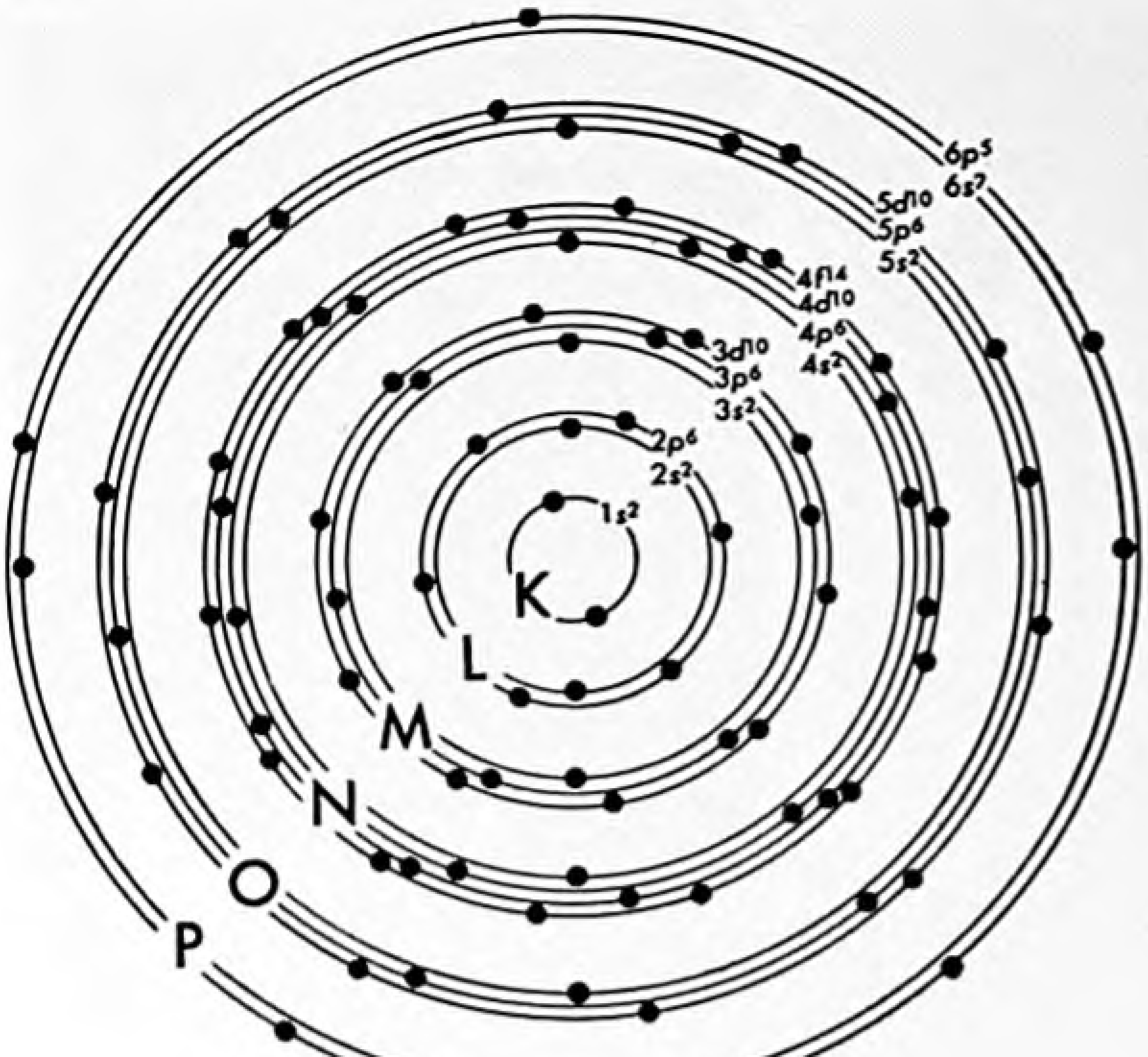
From the behavior of a certain isotope of technetium, Tc^{99}, with a half-life of six hours, it was deduced that this isotope must have another form, or nuclear isomer, with a long half-life. Thus one of the conditions for obtaining macroscopic amounts of the element was fulfilled, namely, that a long-lived isotope must exist. The other condition, a method of preparing the element in quantity, was realized when it was shown that the six-hour Tc^{99} is a fission product and therefore can be made in an atomic pile. The fission yield of technetium is high; Tc^{99}

ONE ELEMENT, the synthetic rare earth promethium, has a characteristic X-ray spectrum with two lines.

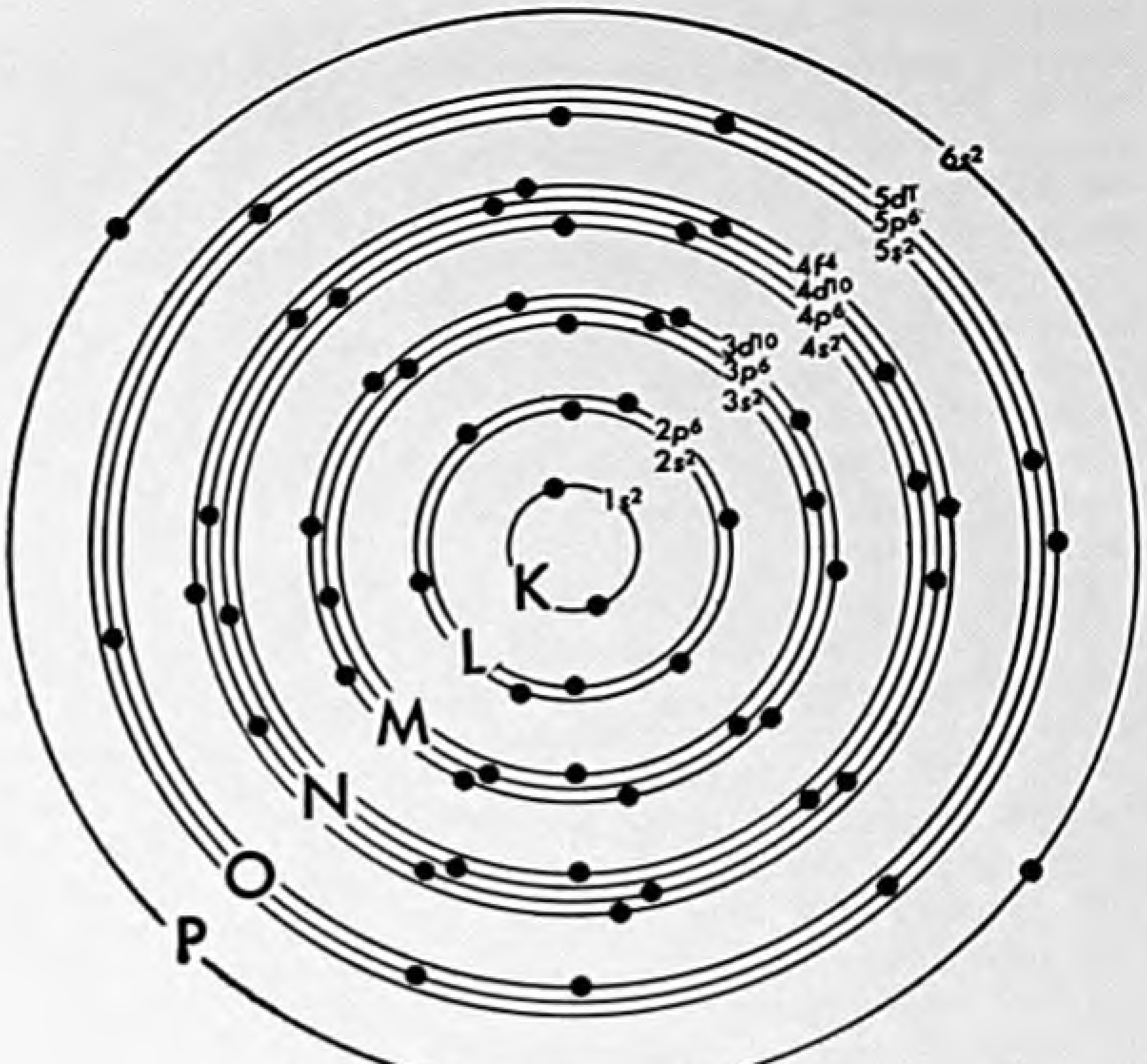
THREE ELEMENTS in succession, neodymium, promethium and samarium, have successive pairs of lines.



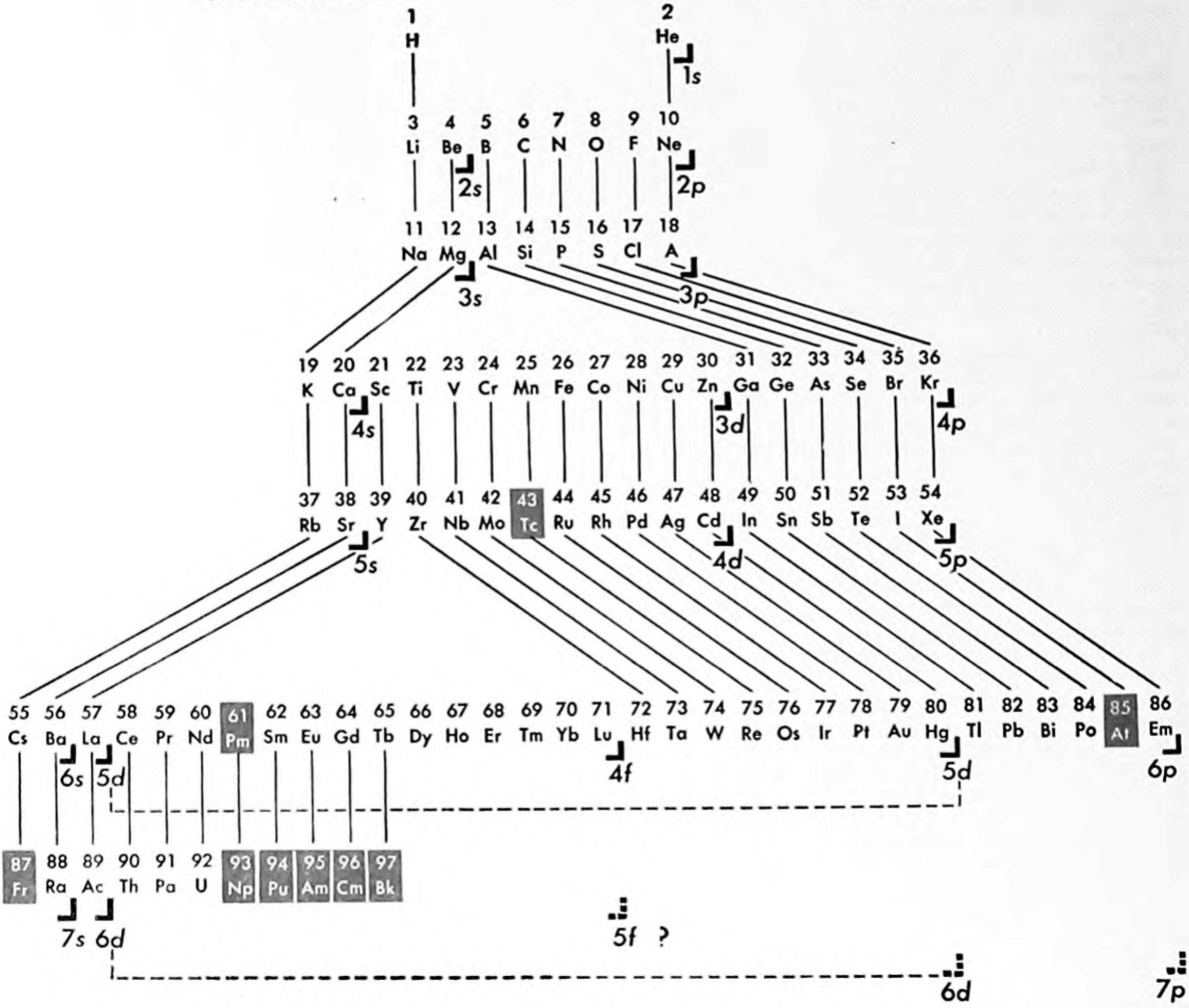
TWO ELEMENTS, neodymium and samarium, have X-ray lines with a gap between them for promethium.



ASTATINE



PROMETHIUM



constitutes 6.2 per cent of the eventual fission products in a pile, and it can be calculated that the fission of one gram of U^{235} produces 26 milligrams of Tc^{99}. The element can also be made by irradiating molybdenum with neutrons to form Mo^{99}, which decays to Tc^{99}; so two methods for production in quantity are available.

When it became possible to produce suitable amounts of long-lived Tc^{99}, it turned out to have a half-life of close to one million years. With this half-life the element becomes relatively easy to handle. A number of milligrams have now been isolated and many of its chemical properties have been investigated. Like the brilliant violet permanganates familiar to chemists, the corresponding pertechnetates are also brightly colored. Technetium has been prepared in the metallic state and combined in a number of compounds; its optical emission spectrum and X-ray spectrum have been recorded; its mass number has been proved to be 99 by means of the mass spectrograph.

It is almost certain that there can be no primordial technetium in nature, since all possible isotopes of the element seem to have half-lives which are far too short. The only processes that come to mind as possible sources for the continuing formation of technetium in nature are (1) the spontaneous fission of U^{238}, a very rare event, and (2) the action of stray neutrons on molybdenum and U^{235}. If we assume that the yield of Tc^{99} from the spontaneous fission of U^{238} is the same as that from the neutron-induced fission of U^{235}, and that the half-life of spontaneous fission is 10^{16} years, then each kilogram of uranium when dug from the ground would contain only a few millionths of one microgram of Tc^{99}. The other mechanisms for producing Tc^{99} would not yield much more.

Element 61

The next synthetic element we shall consider is the rare earth promethium. The rare-earth elements are a group from cerium to lutetium (atomic numbers 58-71 inclusive). They are closely allied in chemical properties to one another and to their prototype, lanthanum (element 57). The rare earths always occur together and have a marked predilection for remaining together under most chemical treatments. The classical method for separating them is repeated recrystallizations, a most laborious and material-consuming process.

When the rare earths were finally lined up according to atomic number, the space for element 61 remained unfilled. Obviously the place to look for element 61 was in rare-earth ores, but the difficulty of separating rare earths from one another was an obstacle. Although a number of claims to its discovery were made, the question of the existence or nonexistence of element 61 in nature was still unsettled when the possibility of preparing it artificially presented itself. Several groups of investigators irradiated neighboring rare earths and produced some new sources of radioactivity, some of which undoubtedly were isotopes of element 61. But which radioactivities belonged to element 61 and which to new isotopes of known rare earths? This question remained unanswered because the methods of fractionating rare earths were of insuperable difficulty and the radioactivities observed had short half-lives.

Two developments in the Manhattan Project during the war conspired to allow the unambiguous discovery of element 61. Foremost was the development of methods for separating the rare earths by means of synthetic ion-exchange resins. It was found that neighboring rare earths could be largely separated in one pass through a glass column filled with resin. Furthermore, the rare earths came off the column in a definite order, inversely as the atomic number, so that it was fairly safe to assume that element 61 would follow samarium, element 62. The second important discovery was that a relatively long-lived isotope of element 61 occurs as a result of uranium fission. The first positive identification of element 61 came in 1945 from the experiments of J. A. Marinsky and L. E. Glendenin at the Clinton (now Oak Ridge National) Laboratory. They suggested that the element be named promethium (Pm) to draw a parallel between mankind's newly acquired nuclear power and the acquisition of fire, which according to Greek mythology was stolen from the gods and given to man by Prometheus.

The longest-lived known isotope of promethium, Pm^{147}, has a half-life of 3.7 years; nevertheless it has been isolated in the pure state. At the Oak Ridge National Laboratory G. W. Parker and P. W. Lantz obtained promethium from a fission-product mixture, and B. H. Ketelle and G. E. Boyd isolated some from neutron-irradiated neodymium.

Element 85

Let us now consider the third of the four gaps in the periodic table, element 85. The synthesis of this element actually came second chronologically, before promethium. Its manufacture presented a somewhat different problem from that of technetium and promethium. Technetium can be made by irradiating molybdenum, the next lowest element, with neutrons or deuterons; promethium similarly can be made from neodymium, the next element below it. But in the case of element 85 the next lowest element, polonium, itself exists only in trace amounts in nature. Hence polonium cannot serve as the starting material. To make element 85 one must start with the element two numbers below it: bismuth. Consequently it is necessary to add not one but two charges, transmuting element 83 to 85. This can be done by irradiating bismuth with accelerated ions of helium, which has two protons. The synthesis was first accomplished by D. R. Corson, K. R. MacKenzie and Segrè at the University of California (to which Segrè had come from Italy). They named element 85 astatine, from the Greek word meaning "unstable." The -ine ending means that the element is a halogen, *i.e.*, a member of the chlorine family.

The particular isotope of astatine first made was At^{211}, an alpha-particle emitter. As a matter of fact, it exhibits a phe-

THE PERIODIC TABLE at the bottom of the opposite page presents the 97 natural and synthetic elements in horizontal rows to show similarities in their chemical properties. Elements of similar chemical properties are connected by the lines running from top to bottom. Above the symbol of each element is its atomic number, *i.e.*, the number of positive charges in the nucleus or the number of electrons bound by them. The nine synthetic elements are indicated in red. In each horizontal row is one or more half-brackets designated 1*s*, 2*s*, 2*p* and so on. Each of these brackets denotes the filling of a shell of electrons—more properly a subshell—in the succession of the elements. The electron shell structures of two synthetic elements are given in the schematic drawings at the top of the page. In X-ray terminology the shells are designated K, L, M, N, O and P. In spectrographic terminology they are designated 1, 2, 3, 4, 5 and 6. The spectrographic subshells are designated *s*, *p*, *d* and *f*. The maximum number of electrons in any *s* subshell is two, in any *p* subshell six, in any *d* subshell 10 and in any *f* subshell 14. The number of electrons in each subshell is indicated by a superscript; for example $6p^5$ in the outermost subshell of astatine (At) indicates that there are five electrons in the *p* subshell of shell 6. In the case of astatine all the subshells are filled except the last. The case of promethium (Pm) and the other rare earths, however, is more complex. In the rare-earth series from cerium (Ce) to lutetium (Lu) the number of 5*d* and 6*s* electrons remains the same; in successive elements electrons are added in the 4*f* subshell. Promethium thus has four 4*f* electrons. The transuranium elements appear to belong to a second rare-earth series.

nomenon known as "branched decay," which means that some atoms break down in one way and some in another, both sequences yielding alpha particles. One part of a sample of astatine emits alpha particles directly, thereby decaying to bismuth 207; the other part first captures an electron and becomes polonium 211, but the latter is extremely short-lived and promptly gives off a very energetic alpha particle. As a result, At^{211} is observed to decay with two alpha particles of widely differing energies. When observed with proper radiation-measuring equipment, this isotope is as distinctive as a cat with two heads.

The nuclear properties of astatine isotopes have been well charted and they lead to the conclusion that probably no species of this element will have a half-life greater than several hours. The chemistry of astatine has been investigated on the tracer scale by the methods of radiochemistry. Its behavior is that of a halogen considerably more electropositive than iodine, just as iodine is more electropositive than the next lightest halogen, bromine. One means of separating astatine from solutions is by electroplating or chemical plating. There is no easy way to prepare astatine in visible amounts, for its longer-lived isotopes can only be made by the use of a particle accelerator such as a cyclotron. Even these have half-lives of only a few hours, so that work with macroscopic quantities will be exceedingly difficult because of the intense radioactivity.

Element 87

The fourth and final gap in the periodic table was element 87. According to its place in the table, this element should be a member of the alkali family, which includes sodium, potassium and cesium. As in the case of the other missing elements, there had been a number of claims to the discovery of element 87 by conventional chemical methods; the substance so identified had been variously called virginium and moldavium. But we now are virtually certain that there can be no stable isotope of element 87, and furthermore the longest-lived known radioactive isotope of the element has a half-life of only about 20 minutes, so it could not have been detected by conventional chemical methods.

Actually element 87 was first discovered unmistakably through its radioactivity, and it was found as a product of the decay of a heavy element. To understand how it was identified, we must examine briefly the various processes by which the elements at the heavy end of the periodic table decay. There are three separate decay series among the natural heavy elements, known respectively as the thorium, uranium and actinium series. The first series starts with thorium, which breaks down by a number of steps through various isotopes of radium, actinium, polonium and other heavy elements until it finally becomes stable lead. The second series starts with uranium and goes through a number of transformations into other isotopes of these elements until it, too, degenerates to stable lead. The third series, starting with actino-uranium, or U^{235}, proceeds through actinium, from which the series derives its name, and finally ends as still another stable isotope of lead.

Soon after the significance of these decay processes became understood, and it was realized that a number of isotopes of elements between uranium and lead should be found as decay products, attempts were made to locate some isotope of element 87 in a decay sequence. Element 87 of course stands above lead, element 82, so it should be found somewhere in one of these series. It soon became obvious, however, that no isotope of element 87 would result from the known breakdowns in the main pathway of any of the three decay series. But in 1914 Stefan Meyer, V. F. Hess and F. A. Paneth of Austria noted that actinium, Ac^{227}, which was known as a beta-emitter, also decayed occasionally by alpha emission. Since actinium is element 89, its alpha-decay product must be an isotope of element 87, in this case 87^{223}. It was not until 1939, however, that Mlle. M. Perey of France succeeded, by very meticulous radiochemical separations, in obtaining a 21-minute beta-particle emitter which she proved to be the alpha-decay product of Ac^{227}. She later named this new element francium in honor of her native land.

The three natural radioactive series, as we have observed, almost miss element 87 completely. But when the artificial transuranium element neptunium was synthesized, a fourth series, starting with that element, was discovered. And this series yields an isotope of francium, Fr^{221}, in its main decay sequence. This isotope arises from the alpha decay of Ac^{225}. It has a half-life of only five minutes, decaying by alpha emission to an astatine isotope of .02-second half-life. Subsequently it was found that Fr^{221} is also obtained as a decay product in a sequence starting from the artificial isotope thorium 233 [*see page 353*].

The short half-lives and inaccessibility of the francium isotopes have discouraged chemical investigation of the element. It appears to behave like an alkali element in solution; one item of note is that francium has great volatility when the solution is evaporated to dryness and brought to a temperature of several hundred degrees. This is a property of alkali elements that begins to be prominent with cesium and is accentuated with francium.

Thus the gaps in the classical periodic system, covering the elements from hydrogen to uranium, are now completely filled. We turn next to the transuranium elements.

Beyond Uranium

The search for transuranium elements, a quest born of scientific curiosity, was destined to be the trigger for a series of events which within a decade were to rock the world and burst upon the consciousness of every literate human being. These events, of course, were the discoveries that led to the exploitation of nuclear energy, in particular as a weapon of mass destruction. Other fundamental scientific discoveries undoubtedly have had equal or greater effect on mankind's mode of existence in the past, but none literally exploded in his face as has this one.

In 1934 Frédéric and Irène Joliot-Curie of France made the exciting observation that an ordinary stable element could be made radioactive by irradiating it with alpha particles of natural origin. This discovery of artificial radioactivity immediately stimulated research toward preparing radioactive forms of many elements. Two other extremely important developments were taking place at about the same time. One was the development by E. O. Lawrence at the University of California of the cyclotron, which was soon able to accelerate charged particles to energies far beyond those of naturally occurring alpha particles. This discovery made it possible to bombard and transmute the heavier elements for the first time, for alpha particles from natural sources can penetrate the nuclei of only the lightest elements. The second development was the discovery by James Chadwick of England of the neutron, an uncharged particle capable of entering any nucleus easily. Neutrons will literally fall into any nuclei at which they are directed. Since neutrons could be prepared by directing radium alpha-particles at a light element such as beryllium, it became possible for anyone who could acquire 100 milligrams or so of radium to produce and study the transmutation of elements. Most prominent in such studies was a group working with Enrico Fermi of Italy. They soon found a means of preparing transuranium elements by making use of the great avidity of nuclei for neutrons.

It was already known that if the heaviest stable isotope of an element captured a neutron, the nucleus became unstable and decayed to the next higher element by beta emission; this method, as we have seen, can be used to produce technetium from molybdenum and promethium from neodymium. Suppose this process were applied to uranium, the heaviest element. U^{238} should capture a neutron and become a heavier isotope, U^{239}, which would be beta-unstable and

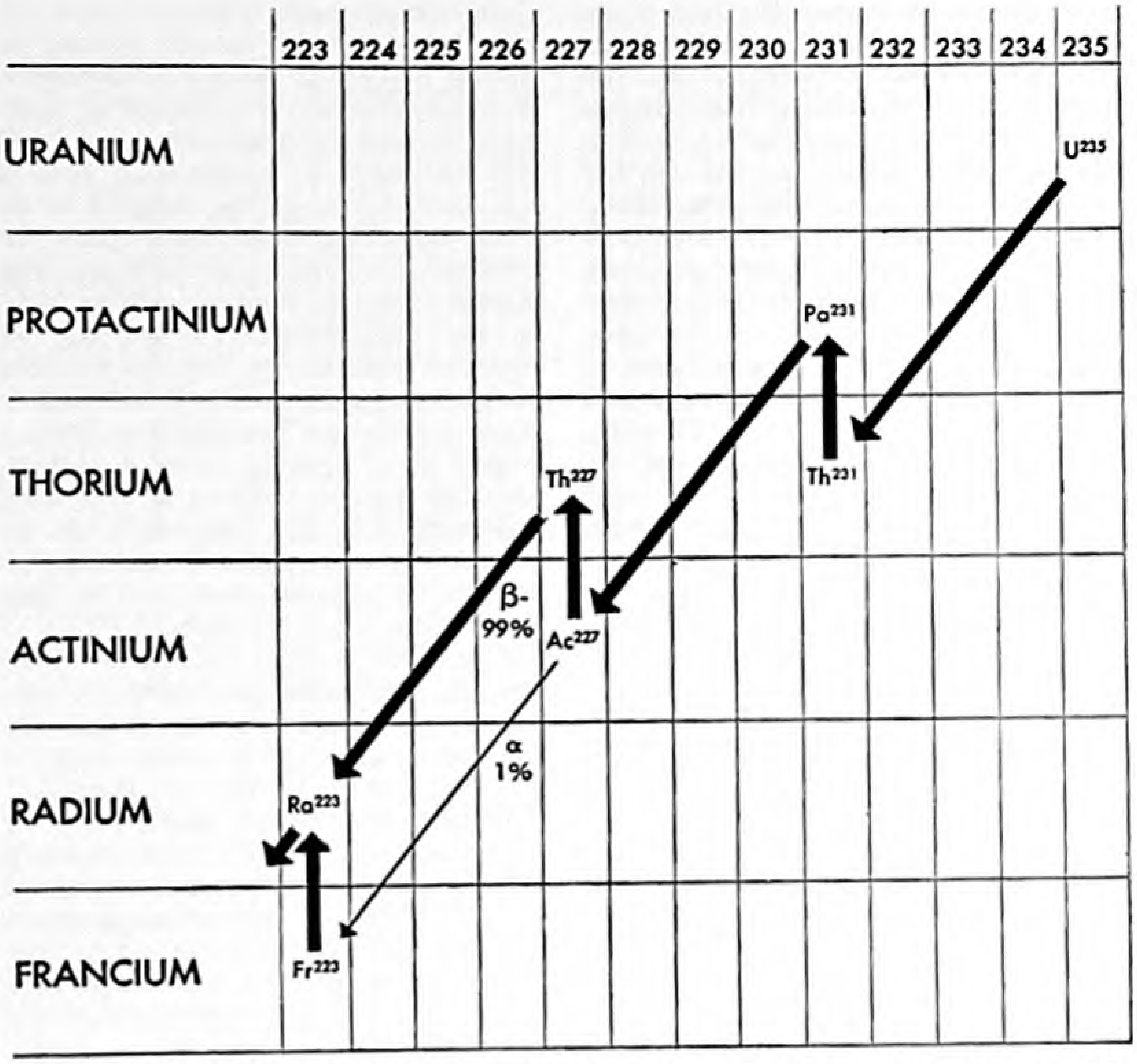
decay to element 93—a brand-new element outside the periodic table! When this experiment was tried, the experimenters experienced a shock: instead of observing just one or two radioactivities from the product, they found a bewildering array of radioactivities. For some time it was thought that these activities must represent a number of new transuranium elements. Not until several years later was it recognized that the activities came from fission products. Thus the discovery of fission was a by-product of the search for transuranium elements.

With poetic justice the actual discovery of the first transuranium element in turn resulted from experiments aimed at understanding the fission process. Several experimenters, including E. M. McMillan of the University of California, measured the energies of the two main fission fragments by observing the distances they traveled from each other as a result of their mutual recoil when the nucleus exploded. McMillan noted that there was another radioactive product of the reaction, with a half-life of 2.3 days, which did not recoil, at least not sufficiently to escape from the thin layer of fissioning uranium. He suspected that this was a product formed by neutron capture, which does not release much energy, rather than by fission. McMillan and P. H. Abelson early in 1940 deduced by chemical means that this product was surely an isotope of element 93, arising by beta decay from U^{239}. The latter had a half-life of 23 minutes. Element 93 was given the name neptunium (Np) because it was beyond uranium, just as the planet Neptune is beyond Uranus.

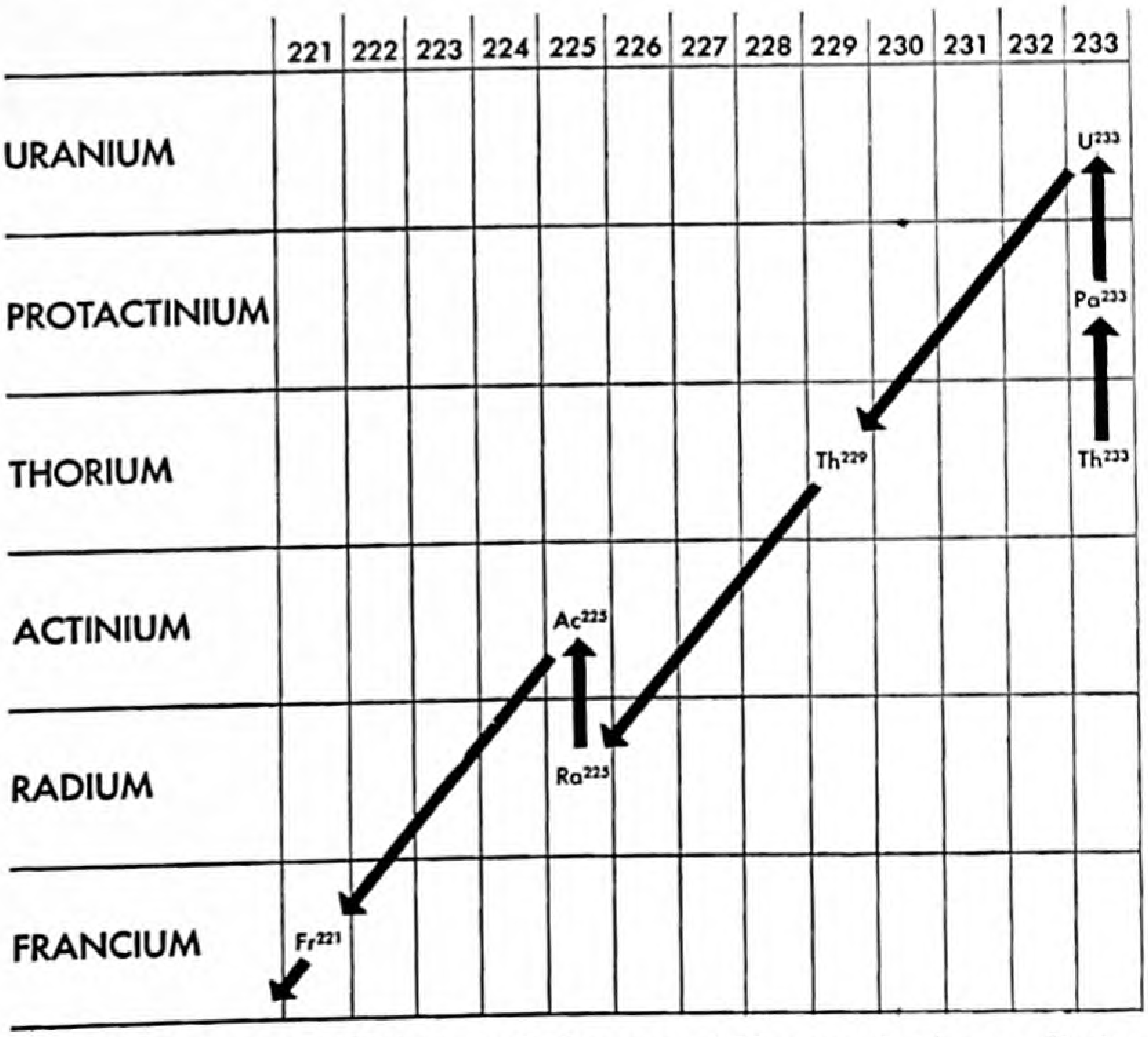
About this time the possibility of a nuclear chain reaction and the production of transuranium elements for military use began to take shape. With the war already in progress, further work on the transuranium elements and related subjects was conducted by physicists and chemists under self-imposed secrecy, at first informally and finally as an organized program.

A neptunium isotope of great practical interest is Np^{237}, discovered in 1942 by A. C. Wahl and Seaborg at the University of California. It is very long-lived (half-life: two million years) and can be made in appreciable amounts as a by-product in the uranium pile. Because it is relatively innocuous, it can be handled experimentally in principle like any normal element.

It was not obvious *a priori* what the electronic configuration and chemical properties of neptunium might be. Uranium was known to have some similarity to tungsten and it was thought that element 93 might be a homologue of the next element above tungsten, rhenium. Yet there was also a possibility that element 93 might be a member of a new



NATURAL FRANCIUM is the result of rare branched decay of actinium 227. Ninety-nine per cent of Ac^{227} decays by beta emission to thorium 227. An almost negligible amount decays by alpha emission to francium 223.



SYNTHETIC FRANCIUM is made in appreciable quantity by irradiating thorium 232 with neutrons. The irradiation forms Th^{233}, the starting point of this table. The decay proceeds through five other isotopes to francium 221.

transition series among the heavy elements, similar to the rare-earth group. It turned out that neptunium bears no resemblance to rhenium. It is much more closely allied to its neighboring element uranium. The evidence is mounting that all of the transuranium elements belong to a new transition series analogous to the rare-earth group. The transuranium elements parallel the rare earths in electronic configuration and have some strong resemblances to them in chemical properties. Just as lanthanum is the prototype element for the rare-earth series, so actinium is the prototype for the heavy-element series. Hence the new group may be called the actinide series. The members of this series known in nature—thorium, protactinium and uranium—had not appeared to be related chemically, but when the transuranium elements were studied latent similarities in the whole group began to appear. Because of certain differences in chemical properties the theory that these elements belong in one series is not accepted by all chemists. It is possible to answer the objections on the basis of a detailed analysis of the evidence, but this is not the place for such a discussion.

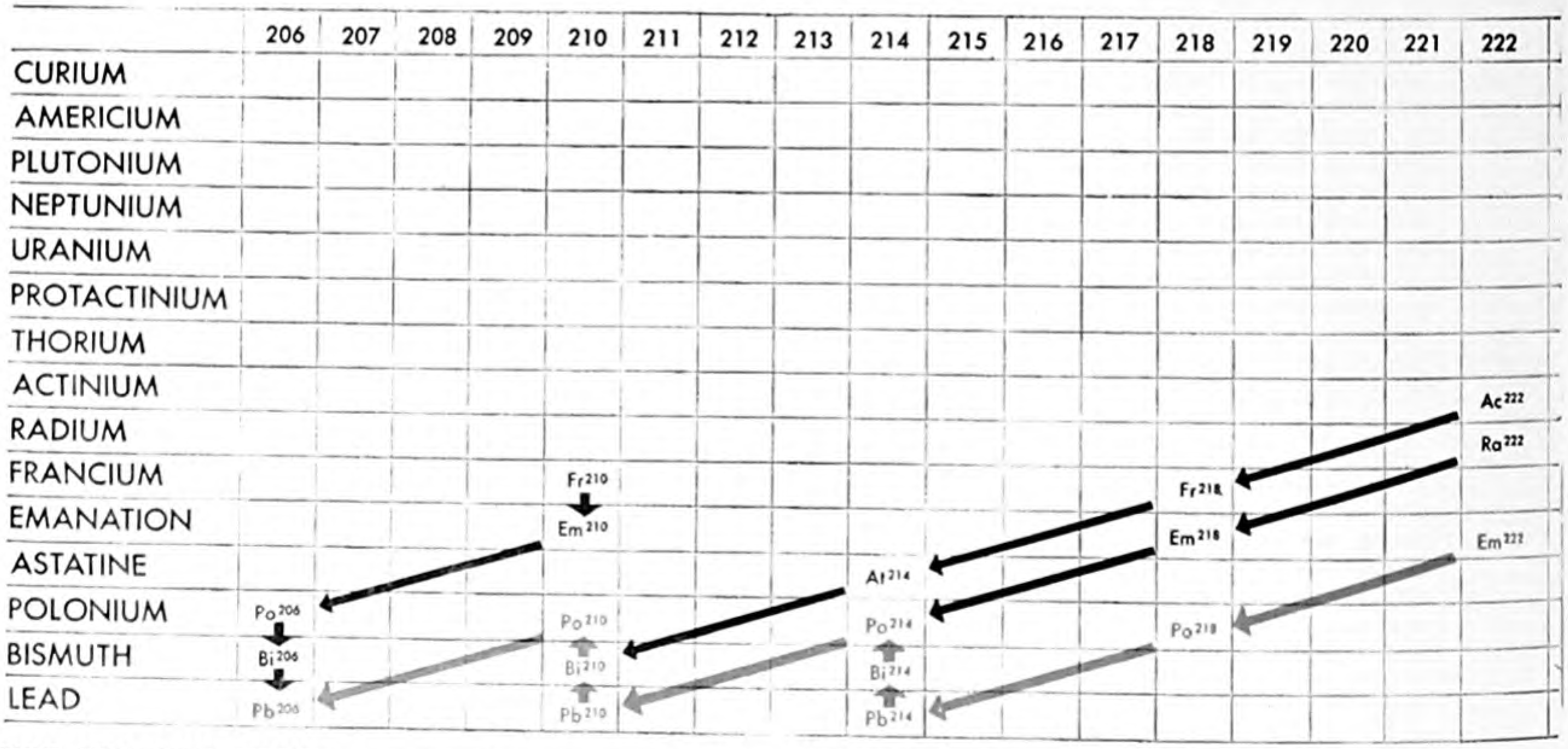
Let us note a few points here, however, on the chemical properties of these heavy elements as a group and their comparison with the rare earths. The rare-earth elements are predominantly trivalent, or in what the chemist now calls the "plus three oxidation state." (Most chemists now use the term "oxidation" to signify the removal or neutralization through bond formation of an element's electrons, whether this is accomplished by the specific method of adding oxygen or by any other means. "Oxidation state" is a somewhat more rigorous term for what we used to call the "valence" of an element.) Only a few rare earths can be induced to assume oxidation states other than the trivalent, and these with difficulty. The heavier elements, notably uranium, neptunium, plutonium and americium, are distinctly multivalent, with the trivalent state becoming progressively more stable along the series. Thus trivalent thorium cannot be obtained in aqueous solution; uranium can be reduced to this state only with difficulty; plutonium can be reduced to it fairly readily; for americium it is the principal state, and for curium it is the only one known.

The effort brought to bear on understanding plutonium chemistry has elevated it to the status of one of the common elements, insofar as knowledge of its chemical properties is concerned. The pronounced multivalent nature of an element such as plutonium makes it of great interest in chemical studies, for this single element affords means of observing most of the phenomena of inorganic chemistry. Plutonium is perhaps unique in having four different oxidation states, which coexist in easily measurable concentrations in aqueous solutions. The color changes from one oxidation state to another afford a fitting visual accompaniment to the curious existence of the many states. Plutonium in its trivalent state in solution is a beautiful pure blue; it changes to green or amber (depending upon solution conditions) when oxidized to the quadrivalent state. The next state, pentavalent plutonium, is colorless; the highest oxidation state, six, is bright yellow. It is unfortunate that plutonium, because of its radioactivity, may never be suitable material for classroom use, for it has superb attributes as a teaching material.

Element 94

After neptunium plutonium was of course the next element discovered; its name derives from the fact that Pluto is the next planet beyond Neptune. The work of McMillan and Abelson had shown that Np^{239} decayed by emission of beta particles; therefore the product should be the next higher element, number 94. However, the new element decayed so slowly that it could not be definitely detected through its radioactivity. By the end of 1940 Seaborg, McMillan, J. W. Kennedy and A. C. Wahl did discover element 94 by a somewhat different approach. By irradiating uranium with deuterons they made a new isotope of neptunium which also decayed to plutonium, but in this case the plutonium was sufficiently short-lived to allow its detection. This isotope of plutonium has proved to be Pu^{238}, with an alpha-decay half-life of 90 years, while that for Pu^{239}, the first isotope made, is 24,000 years.

Armed with the information on the chemistry of the new transuranium elements, Kennedy, Seaborg, Segrè, and Wahl in 1941 were able to identify Pu^{239} from strongly irradiated uranium and were able to prove that Pu^{239} would



THE URANIUM SERIES of radioactive elements is mapped on the basis of atomic number or number of protons in the nucleus (*column of elements at left*) and atomic weight or number of protons and neutrons in the nucleus (*numbers at the top*). In nature (*red symbols and arrows*) the series begins with uranium 238, which decays through 13 other isotopes to stable lead 206. The series has now been enlarged (*black symbols*

undergo fission with slow neutrons. The most intensive cyclotron irradiations ever made were then carried out, with the objective of synthesizing sufficient plutonium to determine certain properties that could best be obtained with visible amounts. In September, 1942, B. B. Cunningham and L. B. Werner, working at the wartime Metallurgical Laboratory of the University of Chicago, isolated a few micrograms of Pu^{239}. Thus plutonium became the first man-made element to be produced in visible quantities.

The realization that plutonium could serve as a nuclear fuel like U^{235}, and that it might be made in quantity in a nuclear chain reactor, resulted in man's first practice of alchemy on a production scale. Plutonium is the only synthetic element yet made in kilogram quantities. The huge plants at Hanford (and presumably other plants considerably west of Hanford) are devoted to this task—the solid embodiment of ideas which sounded utterly fantastic a few years back.

Beyond Plutonium

After plutonium had been produced in quantity, the next higher elements, americium and curium, followed in short order. Based on methods worked out for producing neptunium and plutonium isotopes, principally by cyclotron bombardments of uranium, Seaborg and co-workers discovered elements 95 and 96 during 1944 and early 1945. The speed of discovery of these elements was largely due to the accurate forecast of their chemical properties. Element 96, found by Seaborg, R. A. James and L. O. Morgan, was actually discovered before element 95. It was made by the bombardment of plutonium with alpha particles in a cyclotron. This produced an isotope of mass number 242 with a half-life of half a year. Seaborg, James, and A. Ghiorso later produced element 95 by first preparing Pu^{241}, which decayed by beta emission to an isotope of element 95 with mass number 241 and a half-life of slightly less than 500 years.

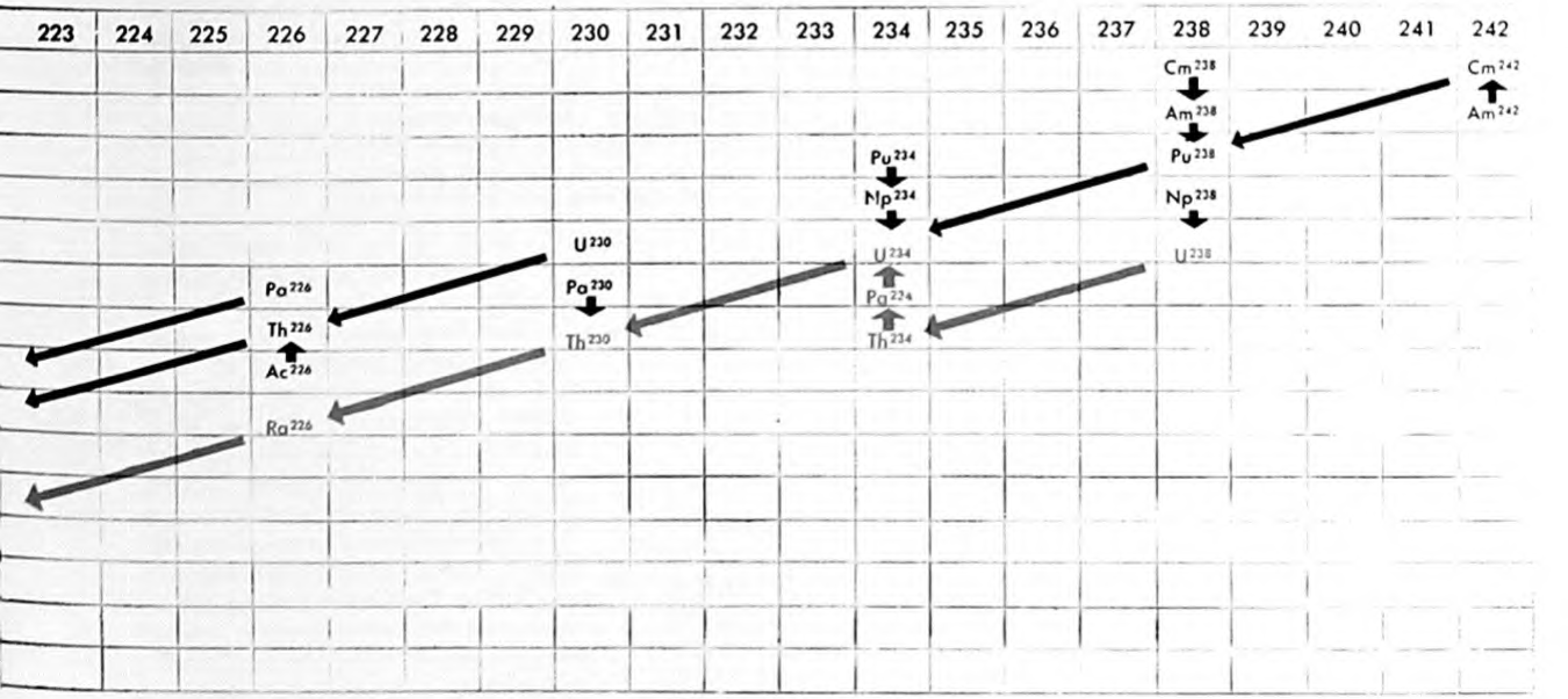
The names for elements 95 and 96 were chosen with regard to their positions in the periodic table according to the actinide concept. The corresponding rare-earth elements are europium and gadolinium, named in the one case for Europe and in the other for J. Gadolin, a Finnish pioneer in rare-earth chemistry. Element 95 was accordingly named americium (Am) for the Americas, and element 96 curium (Cm) in honor of Marie and Pierre Curie.

Both americium and curium have been isolated in a pure state by techniques of ultramicrochemistry. Cunningham was the first to isolate visible amounts of americium. Using some isolated americium, it was possible to convert an appreciable percentage to curium in a pile. From this Werner and Perlman obtained the first curium in the free state. The subsequent work with pure americium and curium has provided abundant evidence for the similarity of these elements to their prototype, actinium.

The elements so far discussed fill in the periodic table completely from hydrogen through curium. If any new elements are to be added, they must lie above curium in atomic number. The beginning of 1950 saw the announcement of the discovery of the first transcurium element, number 97, by S. G. Thompson, Ghiorso and Seaborg at Berkeley. In the naming of element 97, the same convention was used as for the preceding elements. The element in the rare-earth series that corresponds to element 97 is terbium, named for Ytterby, Sweden, where extensive rare-earth deposits were found. It is therefore proper that the new element be called berkelium (Bk), in view of the role played by the University of California at Berkeley in the preparation of most of the synthetic elements.

An isotope of berkelium was first prepared by the irradiation of a minute quantity of americium with cyclotron alpha-particles. Isolation of its radioactivity was accomplished in December, 1949, culminating four years of work on the problem.

Further new elements doubtless can be expected. The difficulty of finding new elements in the transuranium region becomes increasingly severe, however, principally because it becomes less and less likely that isotopes can be prepared with sufficiently long half-lives to allow time for the intricate chemical separations. At what point more basic difficulties will arise cannot yet be said.



***and arrows*) by transmuting natural isotopes into artificial ones with such methods as pile and cyclotron bombardment. These artificial isotopes doubtless existed in nature at one time. Isotopes that decay by emitting an alpha particle lose an atomic number of two and an atomic weight of four. Those that emit a negative beta particle gain an atomic number of one; those that emit a positive beta particle lose an atomic number of one.**

The Authors

ISADORE PERLMAN and GLENN T. SEABORG are leading workers at the University of California's Radiation Laboratory at Berkeley. Between them they have participated in the discovery of all of the transuranium elements. Seaborg, a Nobel prize winner, is director of chemical research at the Radiation Laboratory. He received his Ph.D. from the University of California in 1937 and has been there ever since except for his work at the Metallurgical Laboratory of the University of Chicago during World War II: he was primarily responsible for the development of the chemical separation procedures used in the manufacture of plutonium. Perlman, who is professor of chemistry, is also a California alumnus, having taken his doctorate there in 1940. His wartime career took him to Chicago with Seaborg and then Oak Ridge and the Hanford Plutonium Works. He returned to California in 1945.

Bibliography

THE EIGHT NEW SYNTHETIC ELEMENTS. Glenn T. Seaborg in *American Scientist*, Vol. 36, No. 3, page 361; July, 1948.

DISCOVERY OF THE ELEMENTS. Elvira Weeks. Journal of Chemical Education, 1945.

SCIENTIFIC AMERICAN December 1956, Vol. 195, No. 6, pp. 66-80 OFFPRINT 243

THE SYNTHETIC ELEMENTS II

by Glenn T. Seaborg and Albert Ghiorso

Since 1950, nuclear chemists have synthesized and detected elements 98 through 101. A few short-lived atoms have demonstrated their existence in each case.

An atmosphere of gloom permeated the laboratory that night. In the attempt to produce and identify element 101—the next step in the build-up of a sequence of man-made elements beyond uranium—we had carried out a number of very careful experiments, and all had failed. Now a last experiment was being tried, on the basis of what seemed only a farfetched possibility. At best the minuscule sample of material we had prepared might contain one or two atoms of the elusive 101st element. There was some reason to believe that an atom of element 101 might decay in an hour or two into an atom of element 100, which in turn might break up spontaneously by the fission process. If this barely possible combination of events took place, the creation of element 101 would be signaled in an ionization chamber by a comparatively large pulse of ionization, produced by a fission fragment of its decay product, element 100.

We watched with eyes fixed on a pulse recorder connected to the ionization chamber. An hour went by. The night dragged on toward dawn. The waiting seemed interminable. Then it happened! The recorder pen shot up to mid-scale and dropped back, leaving a neat red line which represented a large ionization pulse—10 times larger than would be produced by an alpha particle. No such pulse had been recorded from natural background radiation in test runs conducted for many days prior to the experiment. It looked highly probable that the pulse was indeed a signal of the hoped-for fission. The vigil continued. An hour or so later the pen recorded a second pulse like the first. We were now confident that we had witnessed the decay of two atoms of element 101—and had added a new member to the roster of chemical elements.

This experiment was probably the most dramatic of the many that have been performed in the University of California Radiation Laboratory in the program of creating new elements. The discovery of element 101 was especially exciting because it was identified on the basis of only a couple of atoms, produced by transmutation in an amount of target material itself so small that it was unweighable. New developments in the sensitivity of techniques had made the discovery possible, and this indeed has been the pattern of additions to the list of elements. Each discovery followed a period of advances in knowledge and methods of analysis. The synthetic elements have been discovered in pairs, so to speak—elements 93 and 94 at nearly the same time, elements 95 and 96 some five or six years later, 97 and 98 four or five years later, and so on. After each pair of discoveries there was a period of mobilization of forces for the next search. The newly discovered elements had to be produced in appreciable amounts for transmutation to the next stage. Techniques and instruments had to be refined.

In a previous article ["The Synthetic Elements," by I. Perlman and Glenn T. Seaborg; SCIENTIFIC AMERICAN Offprint 242] the story of the creation of man-made elements was carried up to element 97. Here we shall deal with discoveries since then—elements 98 to 101—and the growing information about this strange family of heavy elements. The discovery and identification of these elements rests upon the fact that they *are* a family—moreover, a family whose members correspond, in chemical traits, to the members of another family of known and natural elements. This helpful fact [*originally pointed out by Seaborg in 1944 when the chemical properties of the first synthetic elements, neptunium and plutonium, were determined*] makes it possible to predict the chemical properties of each transuranium element before it is actually created. It was the key to the discovery of elements 95 and 96 (americium and curium), and without this clue the elements beyond 96 could not have been discovered at all.

The Actinide Family

The family serving as the guide comprises the so-called rare earths, running from lanthanum to lutetium—elements

EDITOR'S NOTE

Among the many investigators who participated in the work described in this article, in addition to Ghiorso and Seaborg, were S. G. Thompson, K. Street, Jr., G. H. Higgins, G. R. Choppin, B. G. Harvey and G. B. Rossi of the University of California at Berkeley; M. H. Studier, P. R. Fields, S. Fried, H. Diamond, J. F. Mech, G. L. Pyle, J. R. Huizenga, A. Hirsch and W. M. Manning of the Argonne National Laboratory; and C. I. Browne, H. L. Smith and R. W. Spence of the Los Alamos Scientific Laboratory.

57 to 71. The series of heavier elements corresponding to this group begins with actinium (element 89) and includes thorium, protactinium, uranium and the transuranium elements up through the still undiscovered elements 102 and 103 [*see chart on next page*]. Because the first member of the series is actinium, this family of elements are called actinides (as the rare earths are called lanthanides). The family resemblance between the actinides and the lanthanides provides a key to chemical separation and recognition of the individual transuranium elements. A sensitive ion-exchange method has been developed to separate actinides in a mixture: the heaviest transuranium element comes out of the exchange column first and the lighter ones follow it in succession. Thus it is possible to concentrate a newly made element for detection and analysis.

Not only do the heavy elements show a pattern of chemical relationships, but a pattern has also emerged in their physical properties. In the past decade highly sensitive methods have been developed for counting the alpha particles emitted by radioactive atoms and for measuring the energies of these particles accurately. Since each decaying isotope usually radiates its alpha particles predominantly at a single, characteristic energy, the accurate resolution of alpha energies makes it possible to identify isotopes simply by the energy of their alpha emission. Furthermore, the heaviest elements show definite and regular patterns in this emission, so that it is now possible to predict fairly accurately the half-life and alpha-emission energy of a new isotope which one is seeking to synthesize.

To this signature there has been added the identifying marker of spontaneous fission, by which element 101 was discovered. The energy of spontaneous fission does not differ enough from one atom to another to distinguish between them, but the half-life of such fission *is* characteristic for each atom: it varies in a regular manner with increasing atomic weight [*see chart on page 363*]. We were able to predict the spontaneous-fission half-life of the unstable element 100 isotope to which element 101 decayed would be two to three orders of magnitude shorter than that of any known, and so it proved to be.

We also have some knowledge now of the statistical probability of producing new isotopes when a given heavy element is bombarded with particles of a given energy. Needless to say, all these advances in knowledge of the systematic nuclear properties of heavy elements have been very important in the planning of experiments aimed at finding new elements and isotopes.

Elements 97 and 98

We shall begin this account with elements 97 and 98, which were discovered at the end of 1949 and the beginning of 1950. To make them it was necessary to manufacture substantial amounts of the two preceding synthetic elements, 95 and 96. Those elements, americium and curium, are intensely radioactive. In order to detect the new elements to which they were transmuted, extremely efficient methods of separating the new products had to be developed, so that their identifying radioactivity would not be drowned in the radioactivity of the parent elements. The separation of the dangerously radioactive materials required the development of complicated remote-control apparatus.

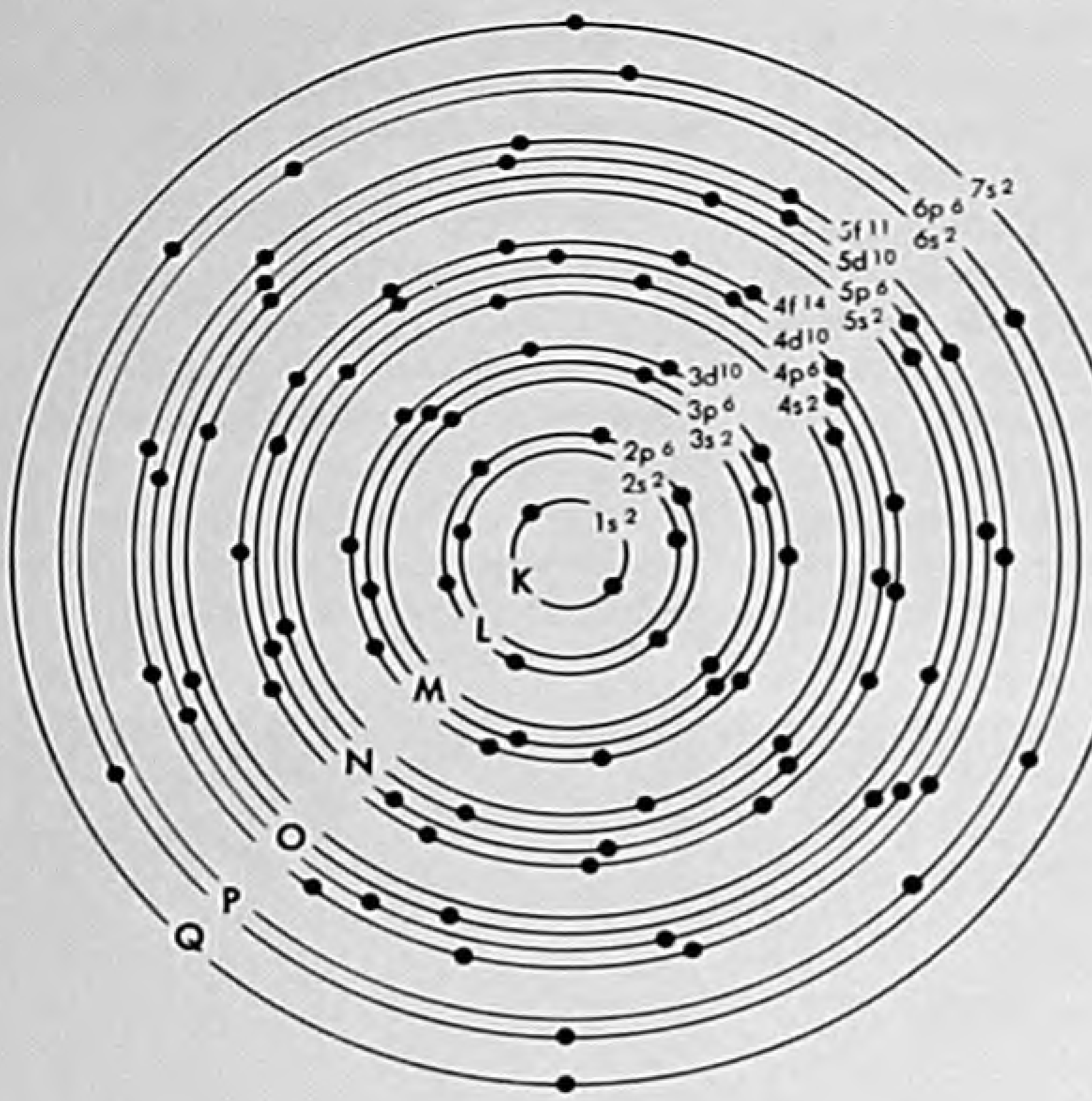
Element 95 (americium) was the starting material. It was made by bombardment of plutonium with neutrons. Intense bombardment of plutonium in reactors over a long period produced milligram amounts of element 95, and this then became the target material for the next steps. Bombarded with alpha particles accelerated to 35 million electron volts in the 60-inch cyclotron at Berkeley, some of the atoms of element 95 were transmuted to element 97. The new element emerged in the form of an isotope which has the mass number 243 (*i.e.*, five more nucleons than uranium 238) and a half-life of 4.6 hours. It decays primarily by the nucleus' capture of electrons from the electronic part of the atom—which is equivalent to emission of positrons by the nucleus.

Element 98, the next step, was produced from element 96, which had been prepared in microgram amounts by bombardment of element 95 with neutrons. The transmutation of 96 to 98, like that of 95 to 97, was accomplished with 35-Mev alpha particles from the 60-inch cyclotron. It yielded an element 98 isotope with the mass number 245 and a half-life of 45 minutes. This new element was identified on the basis of a tiny amount which came to only about 5,000 atoms—less than the number of students at the University of California!

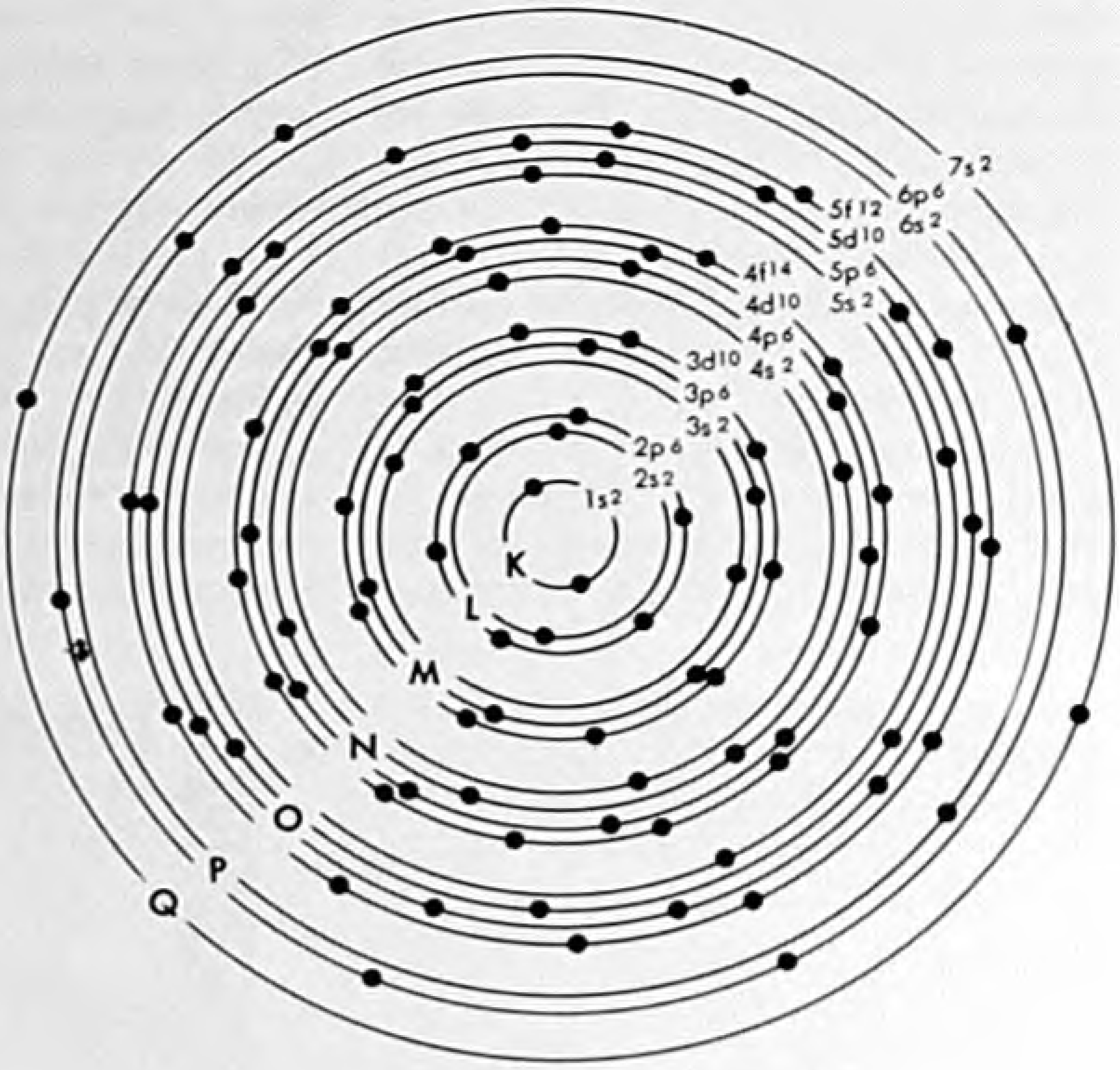
After the discovery of elements 97 and 98, it was found that longer-lived isotopes of them could be made, in substantial quantities, by neutron irradiation of plutonium, americium and curium, which builds up the heavier elements by a series of neutron captures. Eight isotopes of element 97 and 11 of element 98 have now been manufactured, and the long-lived isotopes are available in sufficient amounts for study of the elements' chemical properties on a macroscopic scale. One isotope of element 98 (252) decays in part by spontaneous fission. Since fission always releases some free neutrons, this isotope offers the possibility of use as a tiny, portable source of neutrons for experimental work.

The member of the rare-earth family that is analogous to element 97 is terbium (Tb), named for the town of Ytterby in Sweden, where it was found. Element 97 therefore was given the name berkelium (Bk), after Berkeley. But in naming element 98, whose rare-earth analogue is dysprosium (Dy), the discoverers departed from precedent and

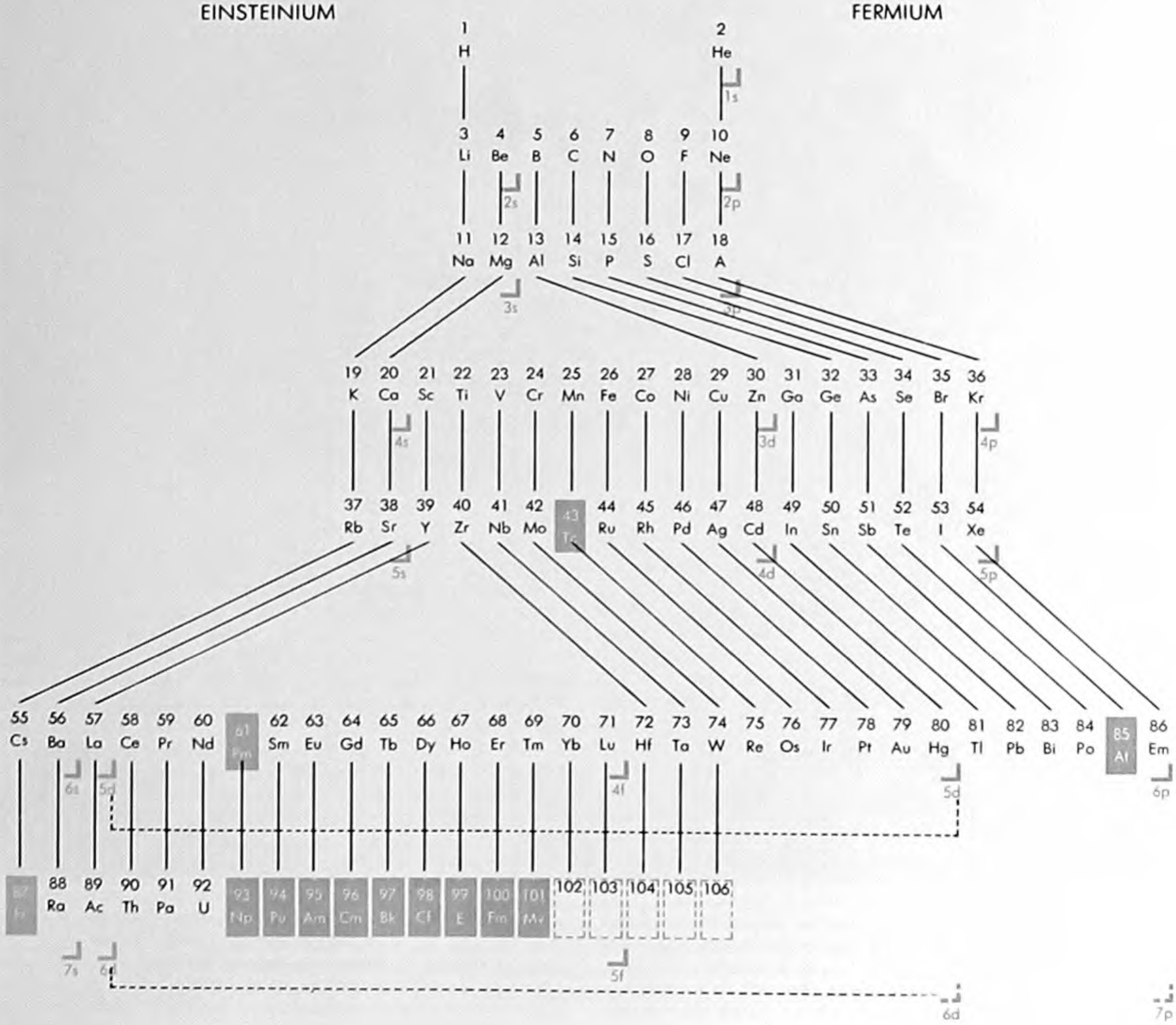
THE PERIODIC TABLE at the bottom of the opposite page presents the 101 natural and synthetic elements in horizontal rows to display the similarities in their chemical properties. Elements of similar chemical properties are connected by the lines running from top to bottom. Above the symbol for each element is its atomic number, *i.e.*, the number of positive charges in its nucleus or the number of electrons bound by them. The 13 synthetic elements are indicated by the solid colored rectangles. The positions of synthetic elements not yet discovered are indicated by the open colored rectangles. In each horizontal row is one or more colored brackets designated 1s, 2s, 2p and so on. Each of these brackets denotes the filling of a shell of electrons—more properly a subshell—in the succession of the elements. The electron-shell structures of two synthetic elements are given in the schematic drawings at the top of the page. In X-ray terminology the shells are designated K, L, M, N, O, P and Q. In spectrographic terminology they are designated 1, 2, 3, 4, 5, 6 and 7. The spectrographic subshells are designated s, p, d and f. The maximum number of electrons (*black dots*) in any s subshell is two, in any p subshell six, in any d subshell 10 and in any f subshell 14. The number of electrons in each subshell is indicated by a superscript number; for example, $7s^2$ in the outermost subshell of einsteinium (E) indicates that there are two electrons in the s subshell of shell 7. In most of the elements all of the inner subshells are filled. The case of the transuranium elements (elements 93 to 101) is more complex. In the rare-earth series from element 58 (cerium) to element 71 (lutetium) the number of 5d and 6s electrons remains approximately the same; in successive elements electrons are added in the 4f subshell. The transuranium elements belong to a series similar to the rare earths.



EINSTEINIUM



FERMIUM

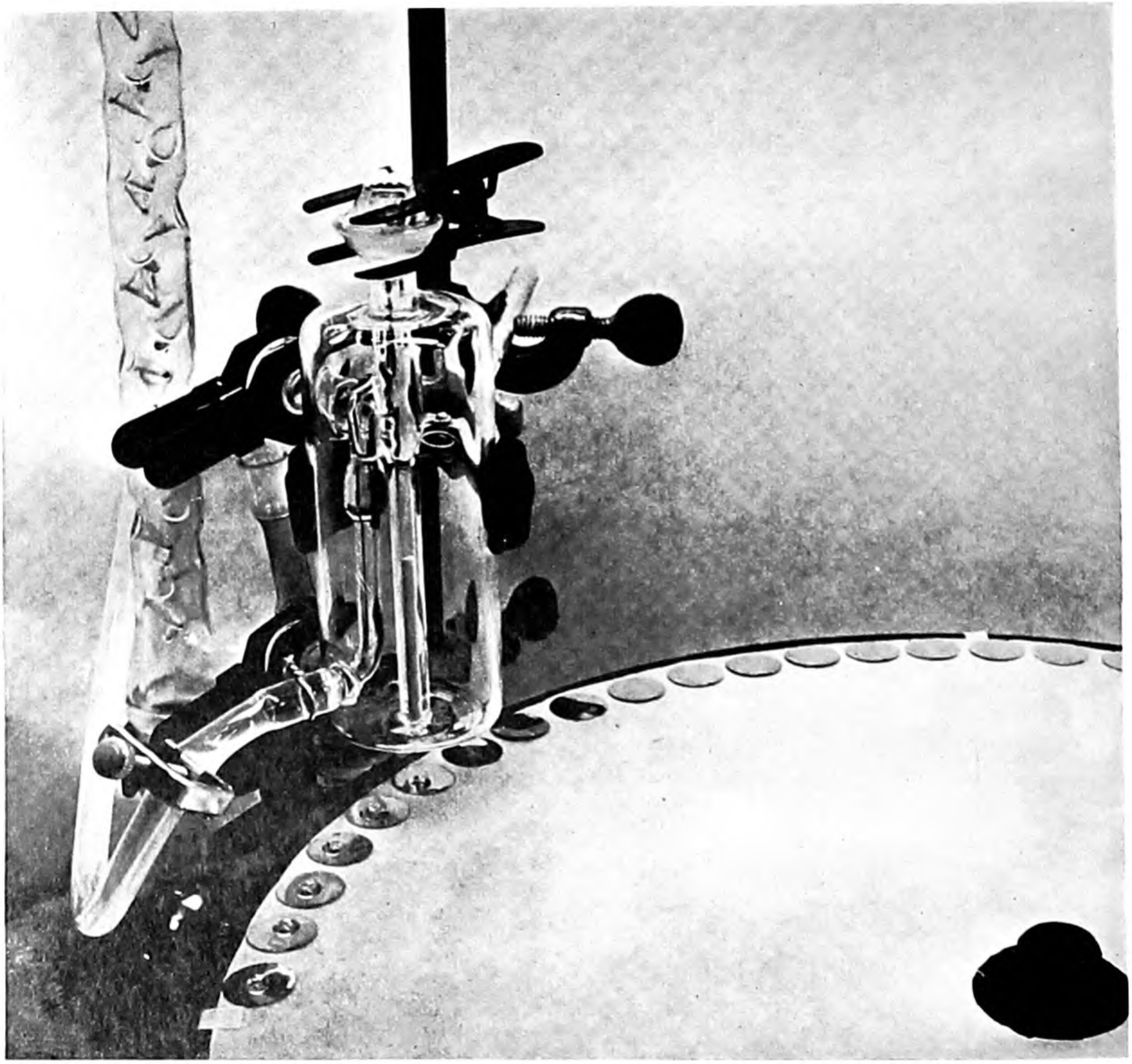


chose the name californium (Cf). In announcing the discovery in *The Physical Review* they remarked: "The best we can do is point out, in recognition of the fact that dysprosium is named on the basis of a Greek word meaning 'difficult to get at,' that the searchers for another element a century ago found it difficult to get to California."

As we have pointed out, the new elements were identified chemically by their behavior in an ion-exchange column. In the drops collected at the bottom of the column the heaviest element of a series comes out first and the following drops carry the other members of the series in order of decreasing atomic number. For any given element, the successive drops have increasing and then decreasing amounts of the substance (the amount being measured by radioactivity). This information is presented in the form of curves on a chart. In two charts shown here [*see next page*] we have plotted this information for the transuranium elements and for their analogues in the rare-earth series. The plots bring out clearly the chemical correspondence between elements in the two groups.

99 and 100

Elements 99 and 100, the next two elements found, represent an outstanding example of unexpected, and incidental, discovery. They were discovered in debris from the test thermonuclear explosion of November, 1952, in the Pa-

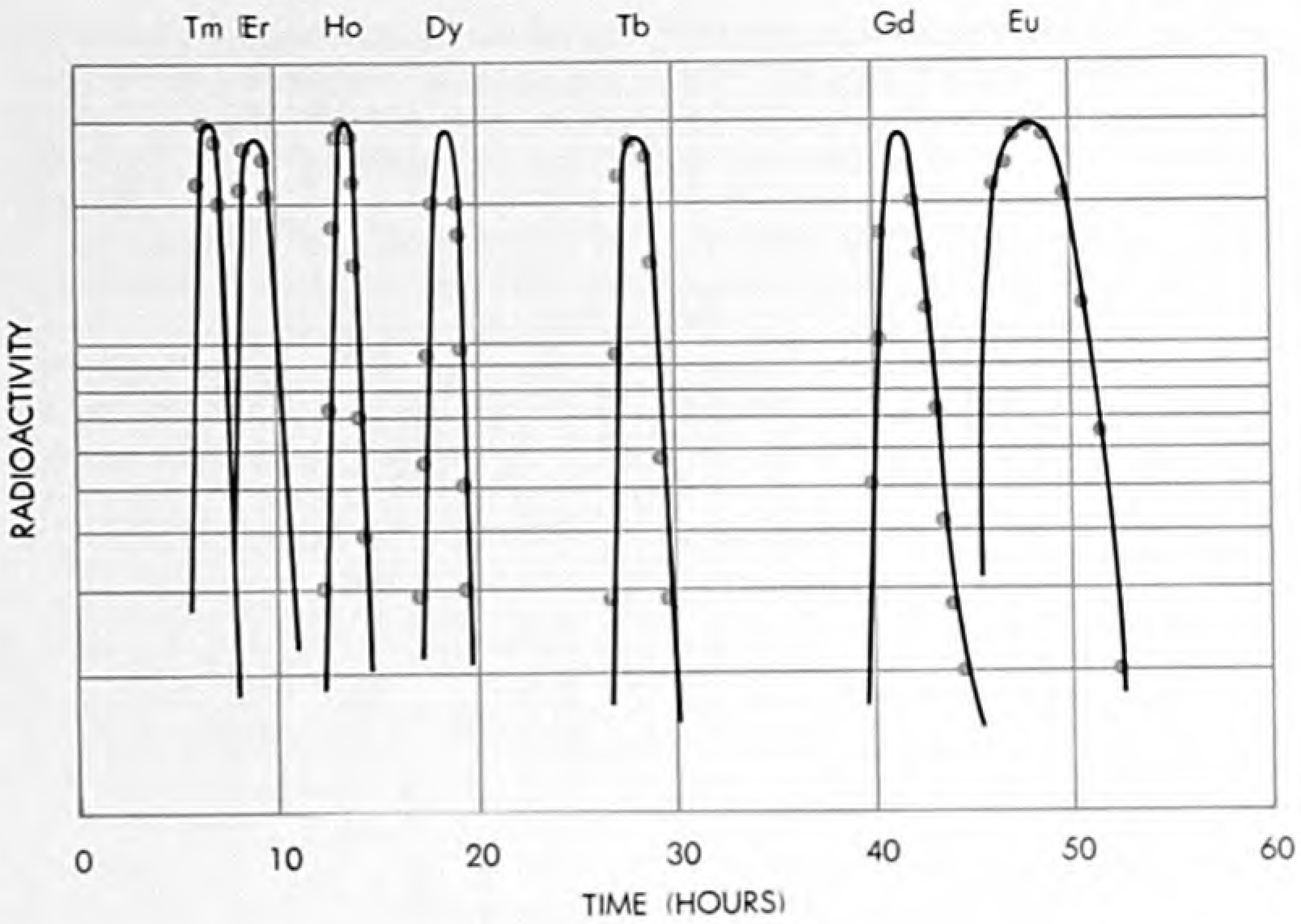


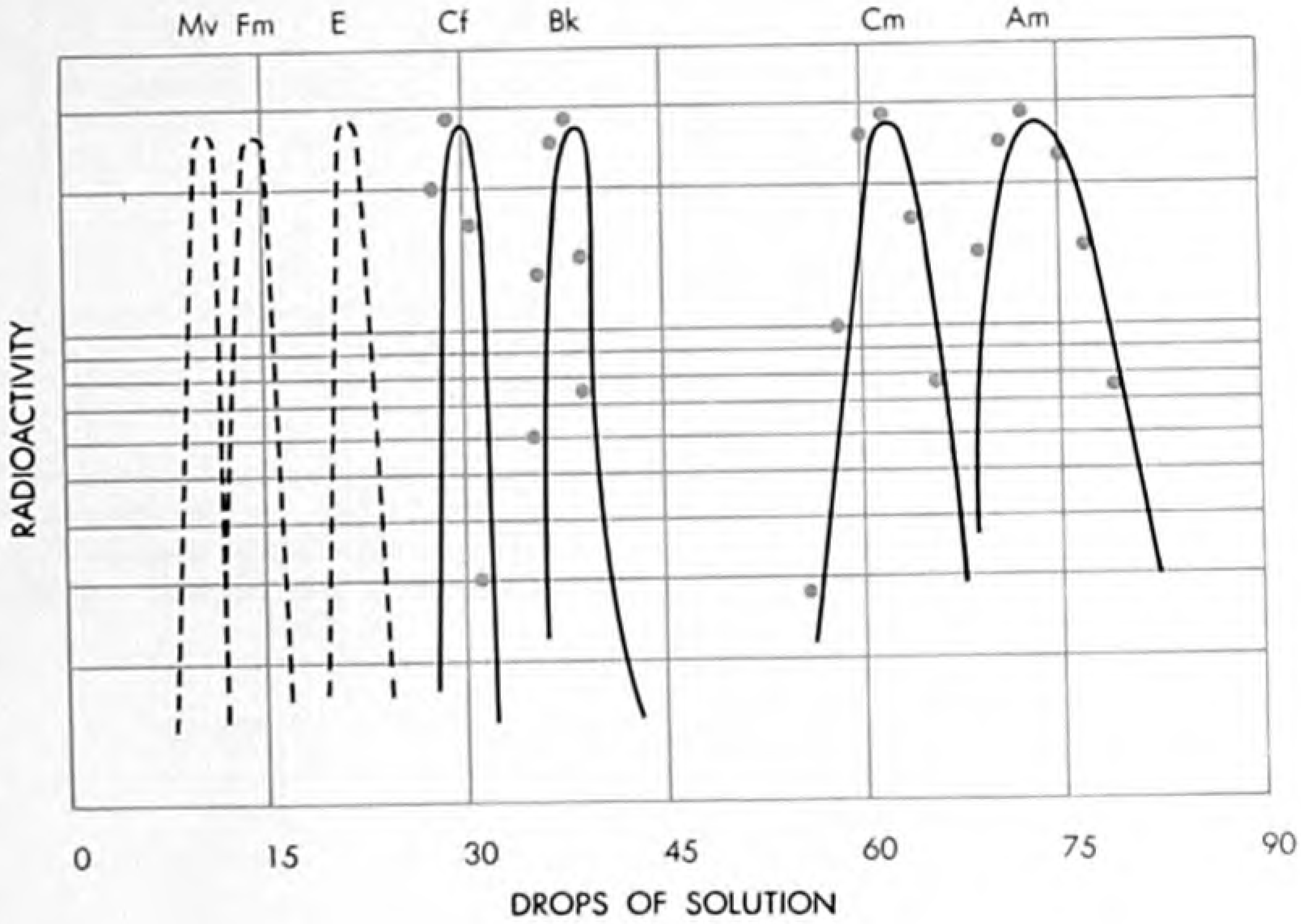
ION-EXCHANGE RESIN is used to fractionate a mixture of transuranium elements. The resin may be seen in the bottom half of the tube contained in the larger vessel at the left. A solution of the mixture is poured into the top of the tube. Each of its constituents passes through the resin at a characteristic rate. Thus the constituents emerge in sequence from the bottom of the tube. They are kept separate by means of the turntable below the tube. As each drop of solution emerges from the tube and falls on a small metal disk, the turntable rotates to bring another disk into position. The drop can now be analyzed by its chemistry and radioactivity.

cific ("Operation Mike"). Material collected on filter paper by drone airplanes flying through the explosion clouds and later in the "fallout" on a neighboring atoll, was brought to a number of U. S. laboratories for chemical investigation. At the Argonne National Laboratory and the Los Alamos Scientific Laboratory it was found to contain some new heavy isotopes of plutonium. This suggested that new elements might have been built up from uranium by many successive captures of neutrons in the explosion, and we undertook at Berkeley to look for elements beyond 98 in the material. Ion-exchange experiments immediately brought one to light. To identify it, more material was called for, and many hundreds of pounds of coral were collected from an atoll near the explosion area and worked over. The material was given the nonsecret code name of "Paydirt"! Paydirt it proved to be, for it led to the positive identification of isotopes of elements 99 and 100 [*see chart on next page*]. The first identification of element 100 was made with only about 200 atoms. The large group of investigators at Berkeley, Argonne and Los Alamos who took part in the work proposed for element 99 the name einsteinium (E), in honor of Albert Einstein, and for element 100 the name fermium (Fm) in honor of the father of the atomic age, Enrico Fermi.

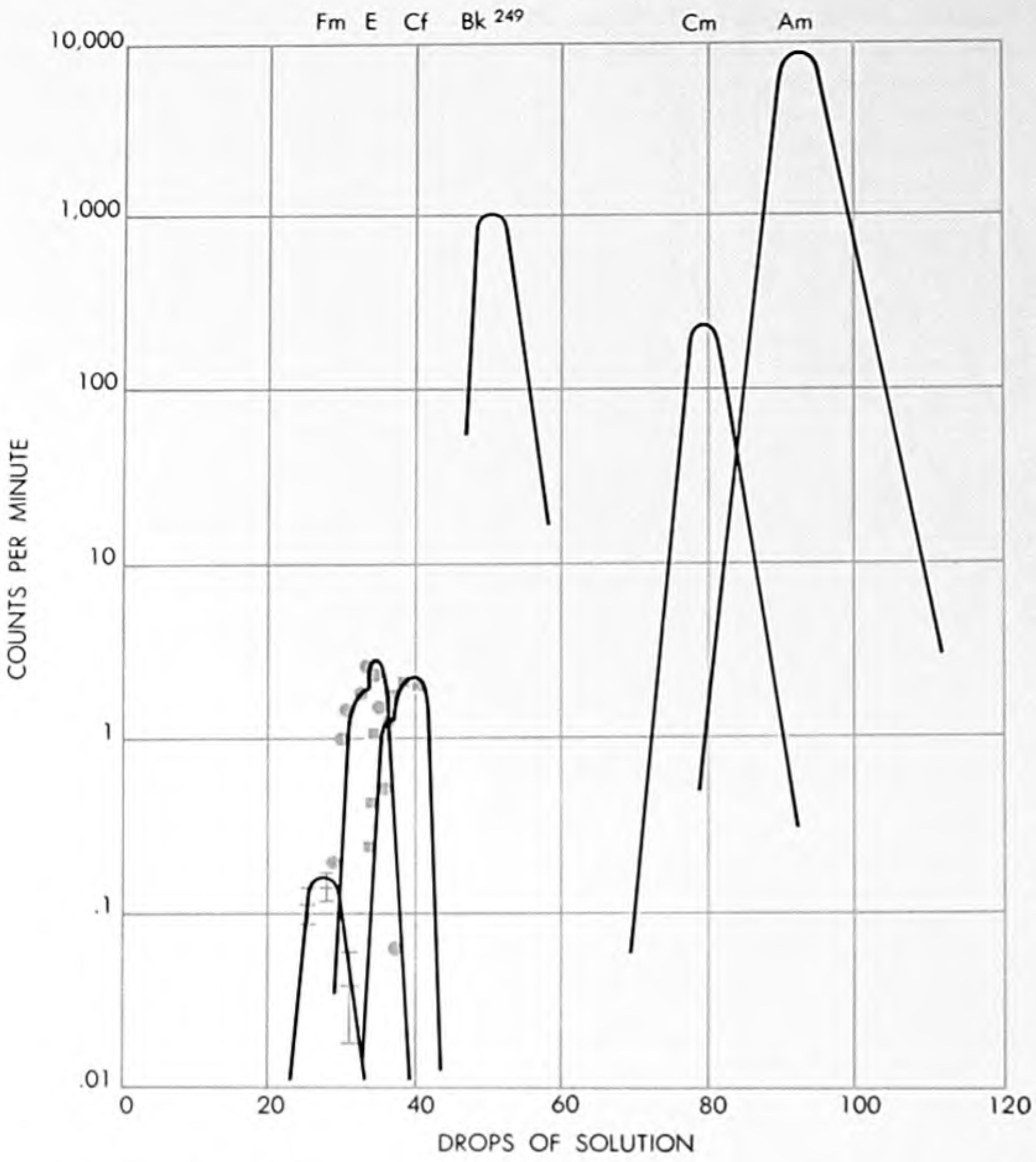
Before these discoveries were declassified and announced to the world, elements 99 and 100 were produced by a number of other methods. Chief among these was prolonged irradiation of plutonium with an extremely high flux of neutrons in the Materials Testing Reactor at Arco, Idaho. The complex series of nuclear reactions (neutron captures and decays) that leads to formation of the elements up to 100, both in the reactor and in the "Mike" explosion, are shown in an accompanying chart [*see page 364*]. To produce these isotopes in a reactor it is necessary to bombard gram quantities of plutonium for two or three years; in the fusion-fission explosion the parent isotopes are created from uranium in a matter of microseconds.

Only tracer amounts of einsteinium and fermium have been available for investigation of their chemical properties. However, there is an isotope of einsteinium which has a half-life of about 270 days, so it should be possible to isolate weighable amounts of this isotope, and substantial quantities of others may be produced by very long and intense neutron bombardment of large amounts of preceding elements. But it appears that

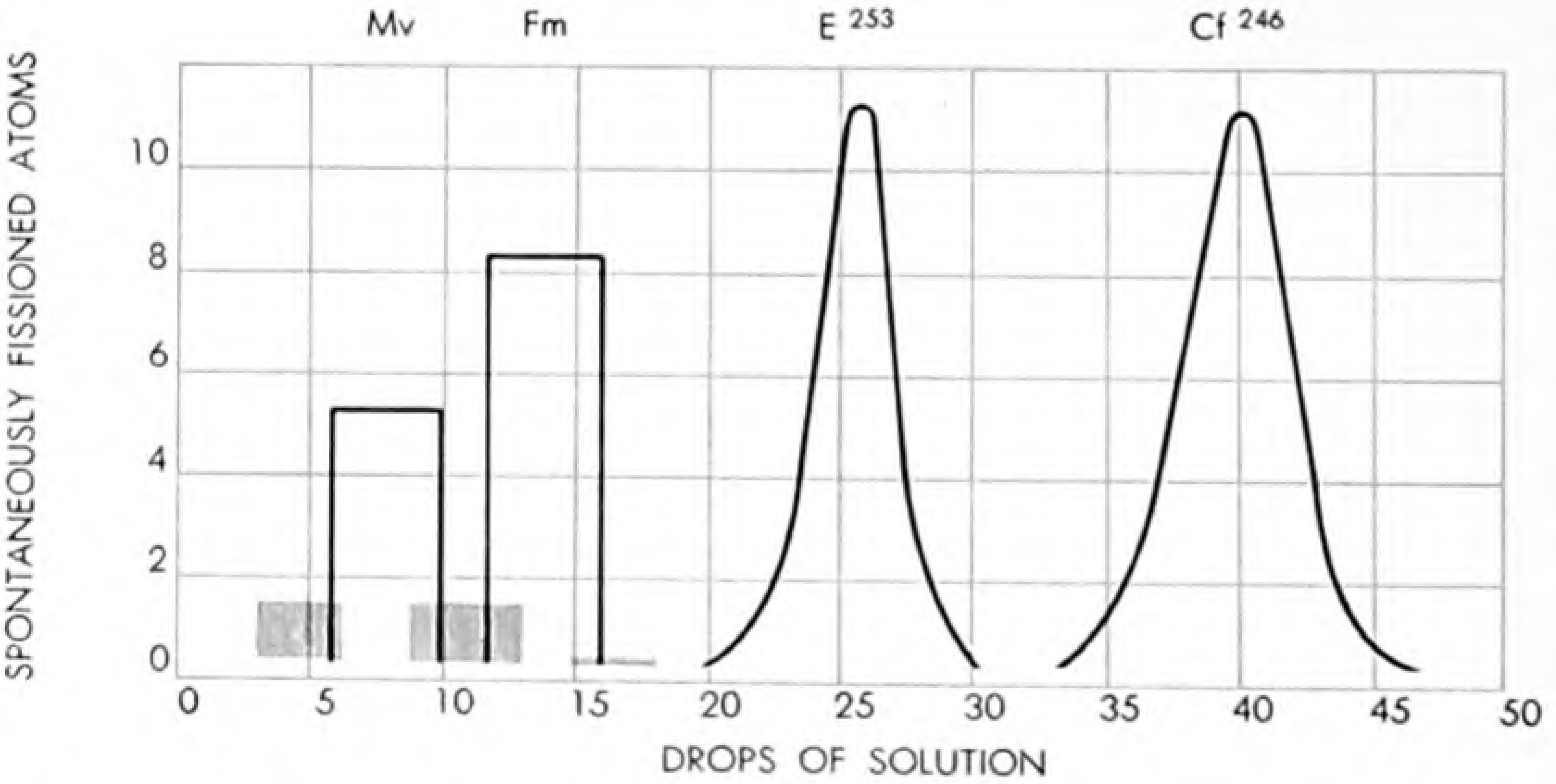




ION-EXCHANGE IDENTIFICATION of californium (Cf) and berkelium (Bk) is presented in the bottom chart. The horizontal coordinate is the number of drops of solution emerging from an ion-exchange column containing a mixture of the transuranium elements californium, berkelium, curium (Cm) and americium (Am). The vertical coordinate is the amount of radioactivity in each sample (*colored dot*). The top chart shows the position of the lanthanide rare earths chemically analogous to the transuranium elements: thulium (Tm), erbium (Er), holmium (Ho), dysprosium (Dy), terbium (Tb), gadolinium (Gd) and europium (Eu). Europium, for example, has chemical properties resembling those of americium, which appears directly below it in the bottom chart. By this analogy the positions of mendelevium (Mv), fermium (Fm) and einsteinium (E) were predicted (*broken curves*). The horizontal coordinate of the top chart (*hours*) differs from that of the bottom chart (*drops of solution*) for the reason that much larger quantities of solution were used to separate the lanthanide rare earths. Actually the coordinates are comparable. The lanthanide rare earths are not naturally radioactive; they were made artificially radioactive by bombardment with neutrons in a reactor to simplify the process of their identification.



EINSTEINIUM AND FERMIUM (E and Fm) were separated by ion exchange from californium (Cf), the berkelium isotope of mass number 249 (Bk-249), curium (Cm) and americium (Am). They were detected by their radioactivity. They could be distinguished in the ionization chamber by the energies of the particles emitted in their radioactive decay.



MENDELEVIUM (Mv) was similarly separated but was detected by the spontaneous fission of its daughter atoms. The open bars indicate the number of fissions in samples emerging from the ion-exchange column in the order expected of mendelevium and fermium (Fm). The gray bars indicate the fission activity of other samples. The curves for einsteinium 253 (E-253) and californium 246 (Cf-246) are based not on fission activity but on alpha-particle activity. The shape of the curves covers measurement of the activity of many samples.

einsteinium is the heaviest element it will be possible to isolate in visible quantities.

Element 101

The discovery of element 101, as we related at the beginning of this article, was in many ways the most difficult and most exciting of all. We decided to make the attempt before there was even a weighable amount of the target material—einsteinium (element 99). The plan of attack was to bombard the isotope einsteinium 253 with the most intense beam of alpha particles achievable by the Berkeley 60-inch cyclotron. All the target material available (from the Arco reactor) amounted to only about a billion atoms. We estimated that the element 101 isotope created would have a half-life of only 10 minutes. It could be calculated that bombardment of the billion atoms of einsteinium with the alpha-particle beam for several hours would yield one detectable atom of the new element! This single atom would have to be separated from the billion atoms of einsteinium and identified by the ion-exchange method in less than 10 minutes.

These requirements indicated a desperate need for new techniques, together with some luck, and fortunately both were forthcoming. The advance in technique was a new method for separating the transmuted element from the target material. The einsteinium target was prepared in the form of an invisibly thin layer electroplated on a gold foil, and the alpha-particle beam bombarded this layer after traversing the foil. Atoms of element 101 produced in the layer would recoil, because of the impact of the alpha particles that effected the transmutations. The recoiling atoms were caught on a second gold foil. This foil, containing the new atoms and relatively free of einsteinium, was dissolved, and the new atoms were then isolated by means of the ion-exchange column.

We first tried to identify element 101 by its decay emission of alpha particles. But we were unable to detect any alpha activity that could be attributed to element 101, even when the time between the end of bombardment and the beginning of the alpha-particle analysis was reduced to five minutes. One of our persistent troubles was that the decay of radon, the rare radioactive gas in the earth's atmosphere, releases alpha particles with about the same energy as that expected for the isotope of element 101. Moreover, the radon decay product in question comes out of an ion exchange column at about the place expected for

element 101! With this type of competition it was very difficult to single out a 101 decay and be sure it was genuine.

Nevertheless, one of these first experiments yielded an event of great significance, as it turned out. Our counter recorded what seemed to be a spontaneous fission! Although we could not be absolutely sure of it, we conjectured that if the event were indeed a spontaneous fission, it might indicate the formation of a new, very short-lived isotope of element 100 from the decay of 101. If that were so, the event might be a means of detecting the existence of 101.

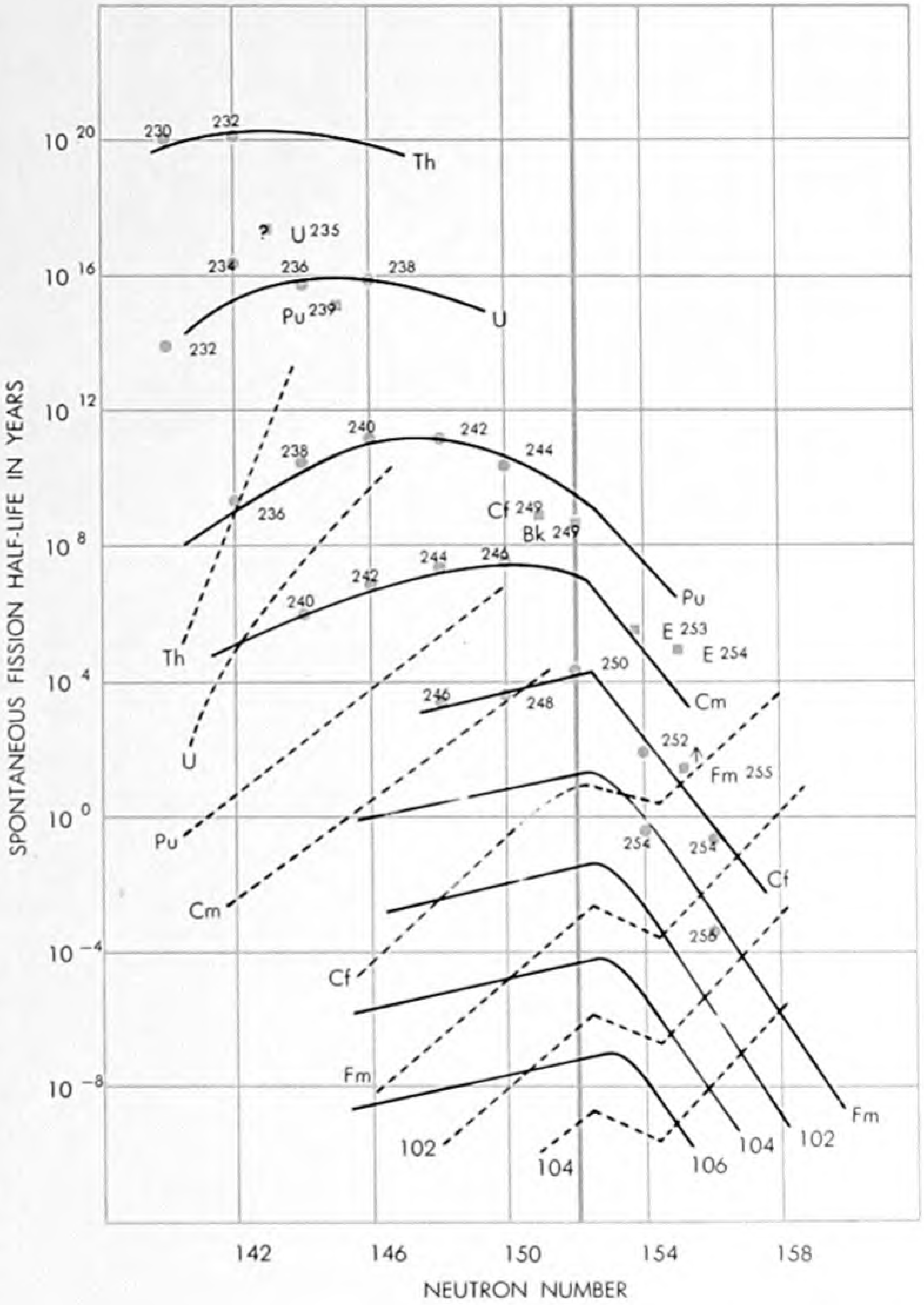
What seemed at the time only a remote possibility turned out to be an actuality. The experiment was revised to look for spontaneous fissions rather than for alpha emissions. The troublesome background interference was now reduced to zero, and the chances of recording fissions in the counter were twice as good as for alpha particles. It turned out that the half-life of the 101 isotope was closer to an hour than to 10 minutes, as we had estimated, but this gain was counteracted by the fact that the probability of formation of the element was considerably smaller than we had thought. The net result was that on the average it was possible to make only one of these new atoms in each experiment: sometimes there were two, sometimes none. A huge fire bell in the hall of the chemistry building was connected to the counting circuit, so that each of these rare events pealed a loud clang of rejoicing. But this sport was put to a justifiable end when it came to the attention of the fire department.

Could the chemical behavior of one or two atoms faithfully reflect the chemistry of the new element? We decided that, under the conditions of our experiments, it actually did so, because each atom went through the same chemical reactions (adsorption and dissolution) perhaps a thousand times during its travel down the ion-exchange column. In other words, we had a substantial statistical sample of its behavior.

The definitive experiments that established the discovery of element 101 were performed in a memorable all-night session. The group bombarded three separate targets of einsteinium for three hours each, and then quickly separated the transmutation products by the ion-exchange method. In each case there was some einsteinium, californium and curium in the mixture in the column, so that it was possible to define the positions in which the elements came off the column. Measuring simultaneously with five counters, we detected five spontaneous fissions in the drops containing element 101 and eight in those in the position of element 100, but none in any other position [*see chart at bottom of page 362*].

By combining the results of all the experiments we deduced that the element 101 isotope has the mass number 256, that it decays by capturing an orbital electron with a half-life of the order of a half-hour, and that the decay product, the element 100 isotope designated as fermium 256, breaks down by spontaneous fission with a half-life of about three hours.

The group taking part in the discovery suggested that element 101 be named mendelevium, in recognition of the great Russian chemist Dmitri Mendeleyev, who was the first to use the periodic system of the elements to predict the chemical properties of undis-

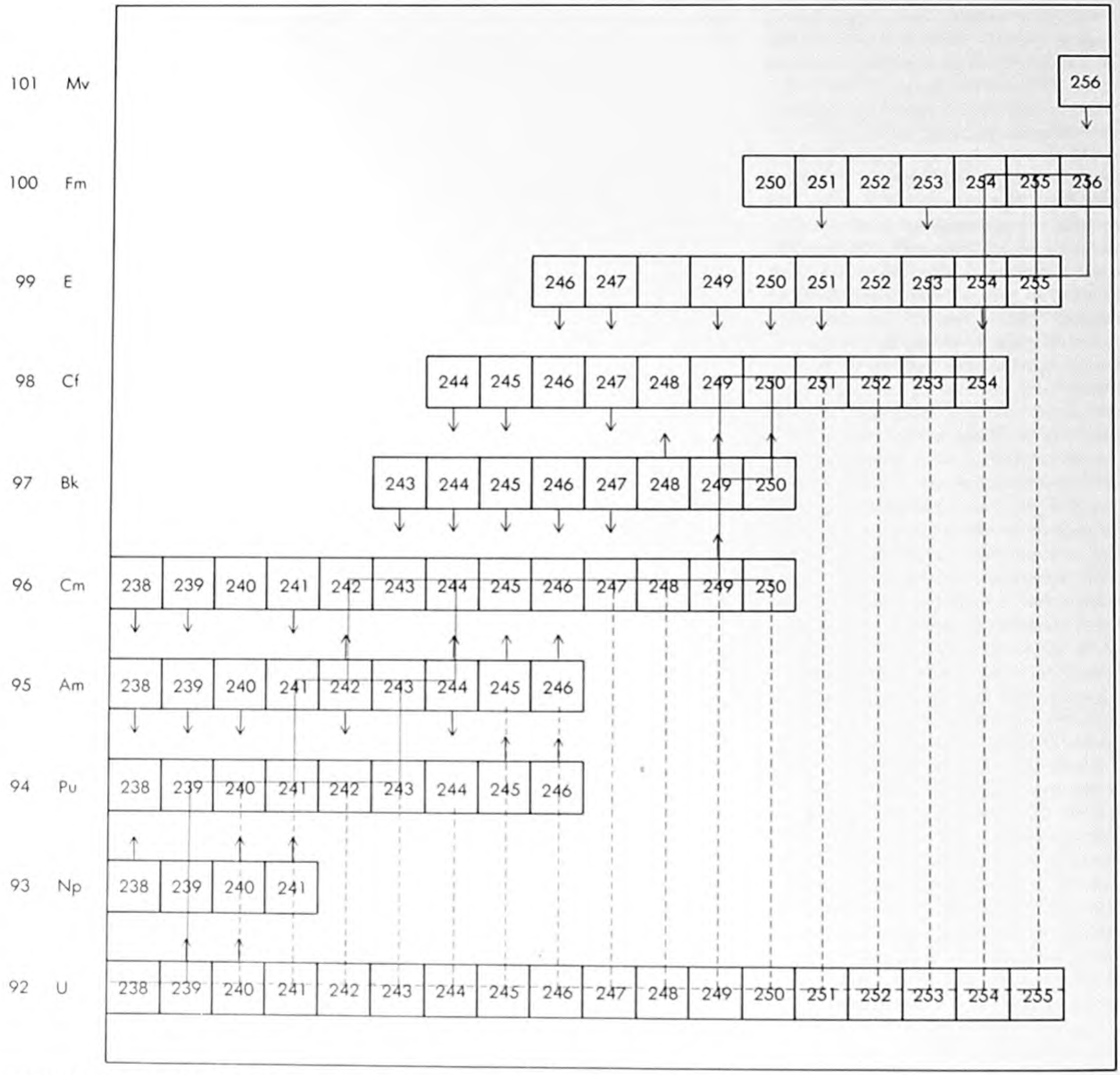


TIME REQUIRED FOR THE SPONTANEOUS FISSION of undiscovered isotopes can be predicted on the basis of the time required for the fission of known isotopes. The round colored dots in this chart indicate the fission half-lives of known isotopes with various numbers of neutrons. The solid curves are based on these points, and suggest where the fission half-lives of undiscovered isotopes will fall. The square colored dots represent anomalous observations. The arrow above the point for fermium 255 (Fm-255) indicates that its fission half-life is at least this long. The broken curves similarly plot the time it takes various elements to decay by emitting an alpha particle. These latter curves are based on actual observations except for the undiscovered elements 102 and 104. The heavy vertical colored line is at 152 neutrons, the point at which fission half-life of isotopes abruptly decreases.

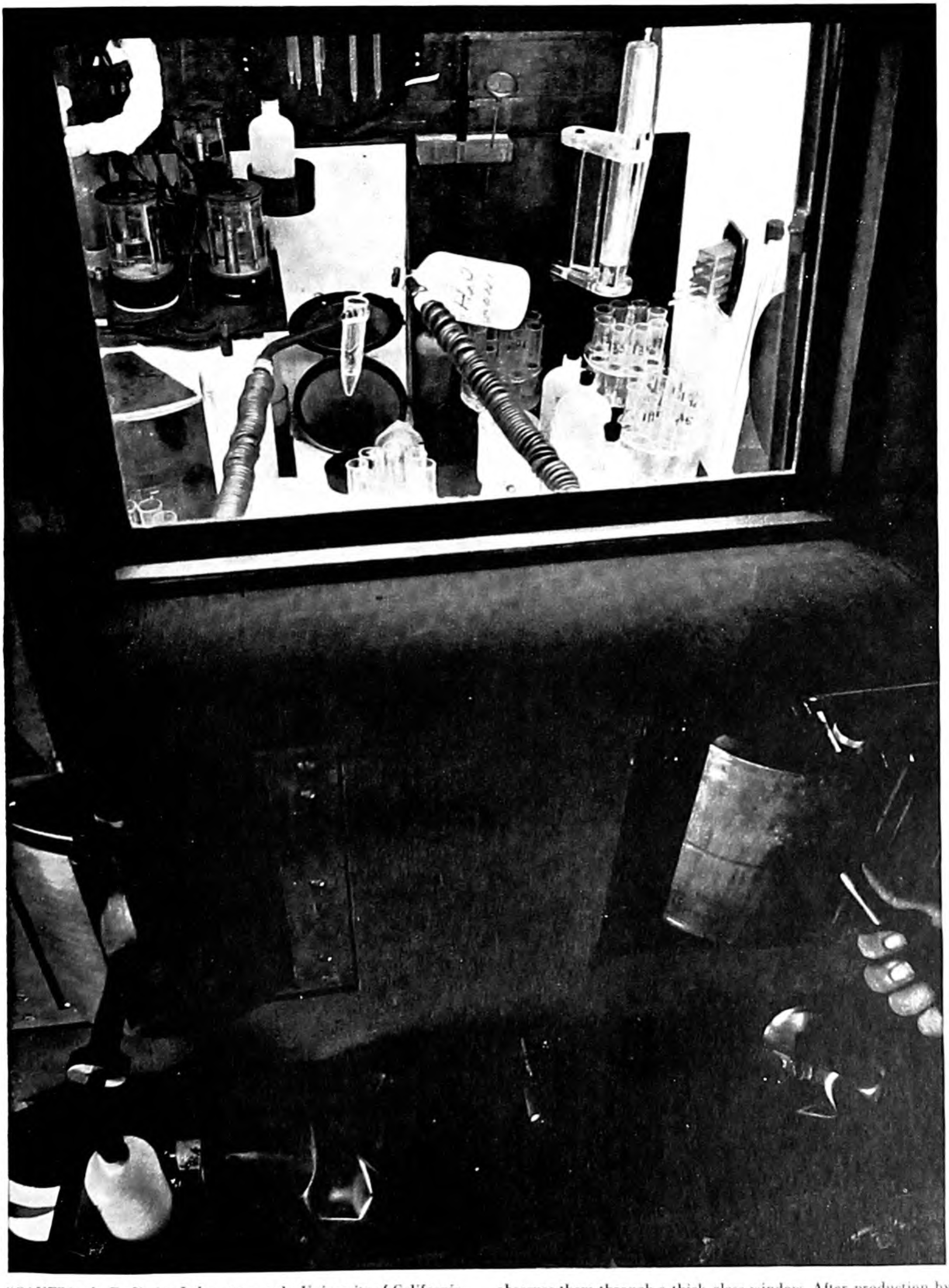
covered elements—a principle which has been the key to the discovery of the last seven transuranium elements.

With larger amounts of einsteinium in the target, it is now possible to produce more than 100 atoms of mendelevium at a time. The decay of mendelevium 256 to fermium 256 has been proved by chemical methods. There is an accumulation of evidence that mendelevium is a typical member of the actinide family, ionizing to the "tripositive" state (*i.e.*, a triple positive charge), and that it has the expected chemical kinship to thulium, its counterpart in the rare-earth family.

We would like to record here a tribute to one of the unsung members of our team, G. Bernard Rossi. His untimely death early in 1956 is a great loss to the Radiation Laboratory. He was chiefly



EINSTEINIUM AND FERMIUM were first made by sequences of nuclear reactions in the "Operation Mike" thermonuclear explosion and later in the Materials Testing Reactor at Arco, Idaho. In this chart the elements are listed vertically according to their atomic number; the isotopes of the elements are listed horizontally according to their mass number, or atomic weight. In the thermonuclear explosion the starting material was ordinary uranium (U-238). By the instantaneous addition of neutrons in the explosion this was built up into a whole sequence of uranium isotopes from U-239 to U-255 (*bottom row*). These rapidly decayed by the emission of negative beta particles into the isotopes above them (*broken colored lines*). In the Materials Testing Reactor the starting material was plutonium (Pu) 239, which had already been made from U-238. By adding two neutrons Pu-239 was transmuted into Pu-241, which decayed into americium (Am) 241. This isotope was in turn built up by the addition of neutrons and so on, following the solid colored lines in the chart. The isotopes of einsteinium and fermium that are not on the pathways were made later by other processes. Some of the pathways in this sequence of reactions are alternative, but they all end at the same point: fermium (Fm) 256. The small black arrows pointing up indicate that these isotopes decay up the scale of elements by the emission of a negative beta particle. The black arrows pointing down indicate that these isotopes decay down the scale by capture of electrons outside the nucleus.



"CAVE" in the Radiation Laboratory at the University of California is used for the chemical manipulation of highly radioactive substances. The chemist performs his operations by remote control and observes them through a thick glass window. After production by intense neutron irradiation of plutonium, the elements berkelium, californium, einsteinium and fermium were isolated in this way.

responsible for the more or less continuous modification of the 60-inch cyclotron which has been necessary for the successful completion of so many of our group's research efforts. For the mendelevium work he improved the operation of the 60-inch cyclotron so as to obtain a useful high-density beam. The successful experiments were due in large measure to his accomplishment of this task.

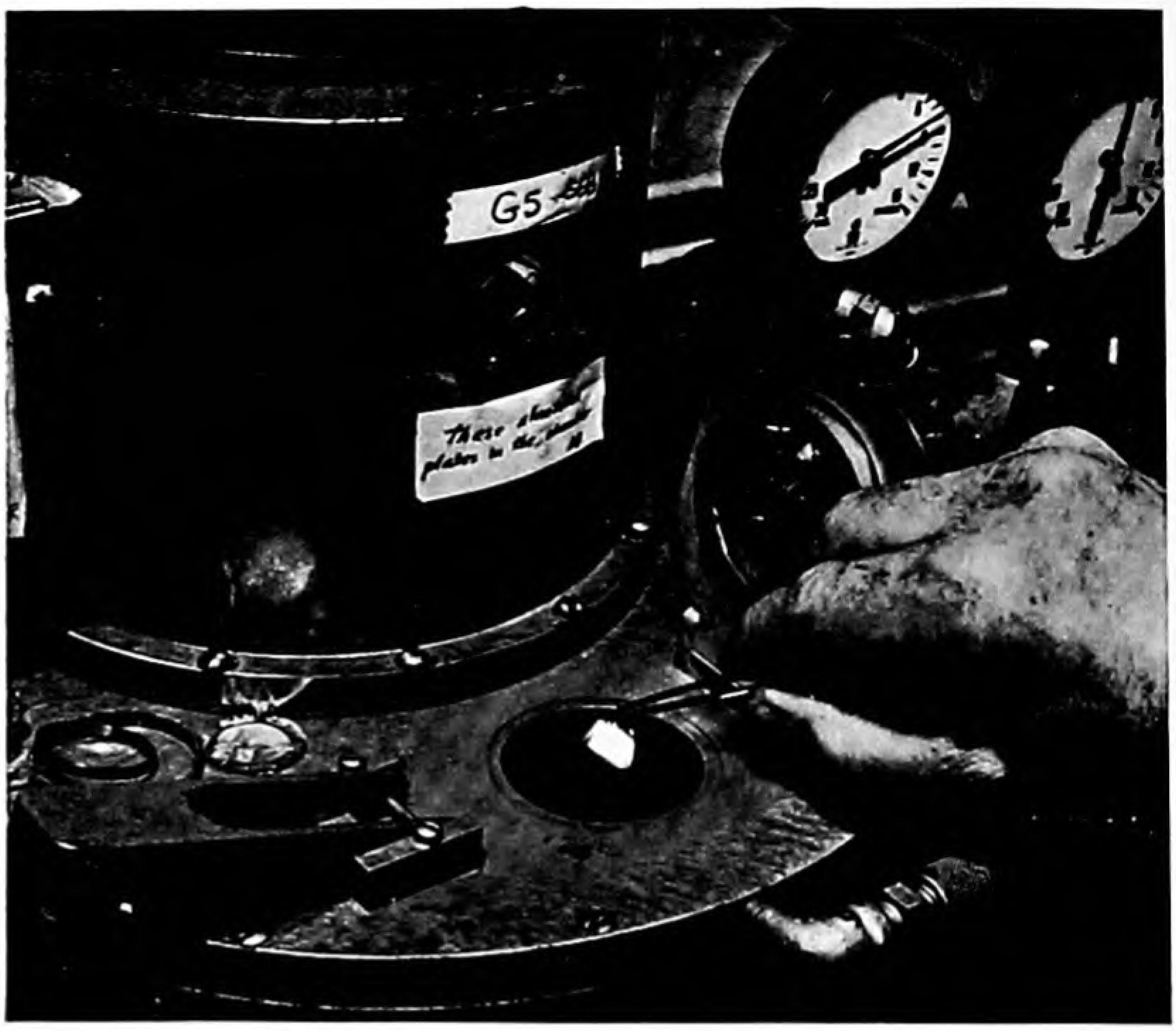
Beyond Mendelevium

In the 15 years since the discovery of the first transuranium element, about six dozen radioactive isotopes of new elements beyond uranium have been made by man. From this work nuclear scientists have learned so much about the radioactive decay of heavy elements that they can now generally predict the decay properties of new isotopes before their discovery. They have also learned that among the very heaviest elements, beginning with fermium, decay by spontaneous fission begins to become about as common as decay by emission of alpha particles. They have found that for both alpha and spontaneous fission decay, predictable regularities are most easily discernible in nuclei with an even number of protons and an even number of neutrons. Isotopes with an odd number of protons or an odd number of neutrons or an odd number of both have slower rates of decay than do those of the regular, even type.

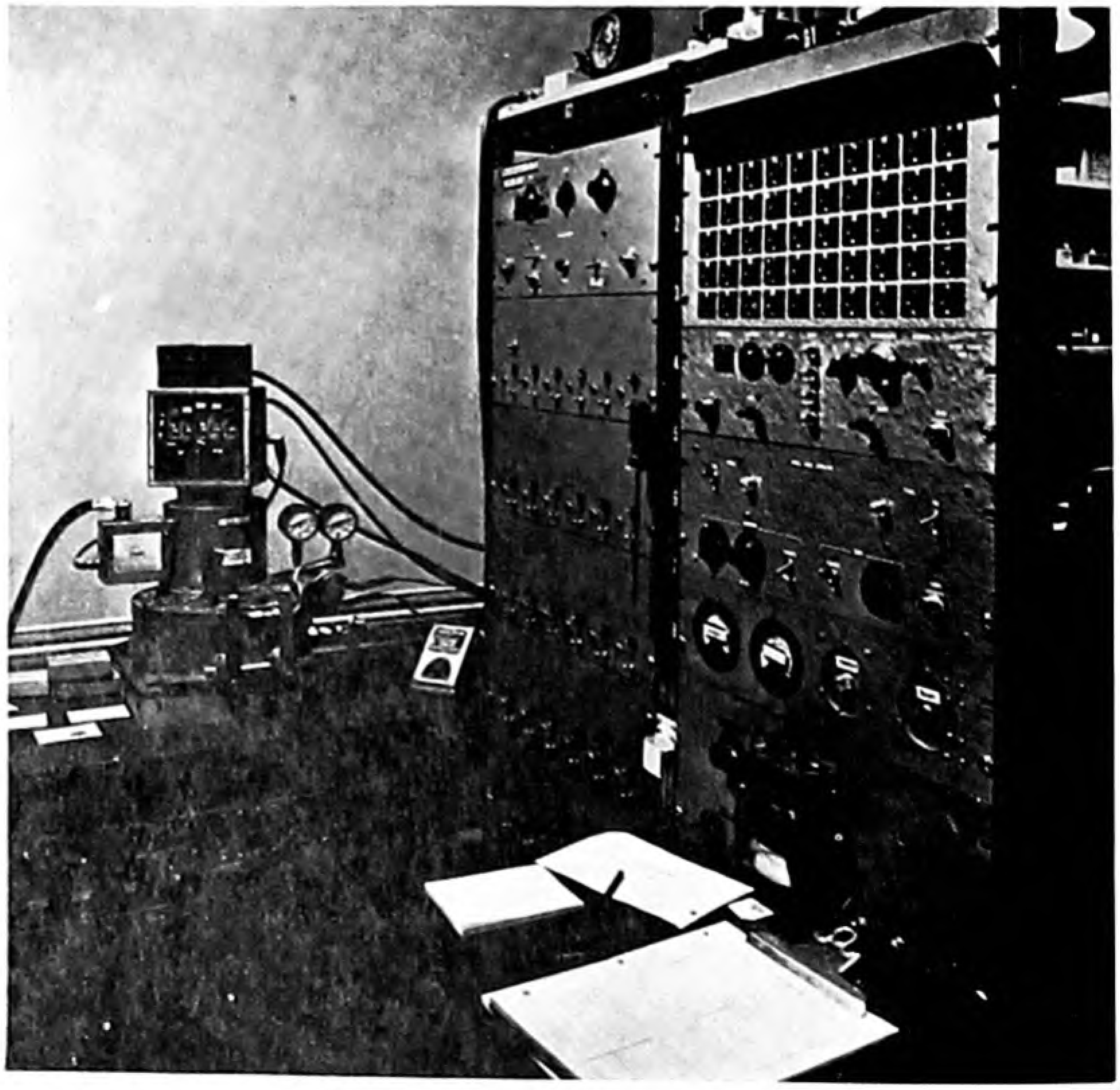
With these considerations in mind, let us have a look at what the future may hold. Unfortunately the half-lives seem to become shorter and shorter with increasing atomic number. By the time elements 104 and 105 are reached, we shall probably find that the longest-lived isotopes that can be made will exist barely long enough to enable chemical identification to be made. It is likely that thereafter we shall have to rely entirely on predicted decay properties, rather than on chemical identification, for the discovery of any further elements. Careful measurements of these properties should allow us to extend the list up to about element 108.

How may these heavy elements be synthesized? The method of build-up by multiple neutron additions seems to hold little promise for making elements beyond fermium, because some of the necessary steps are too short-lived. For example, the isotope fermium 258 will probably have a half-life of only about one minute. It will not accumulate to sufficient concentration to continue the build-up sequence.

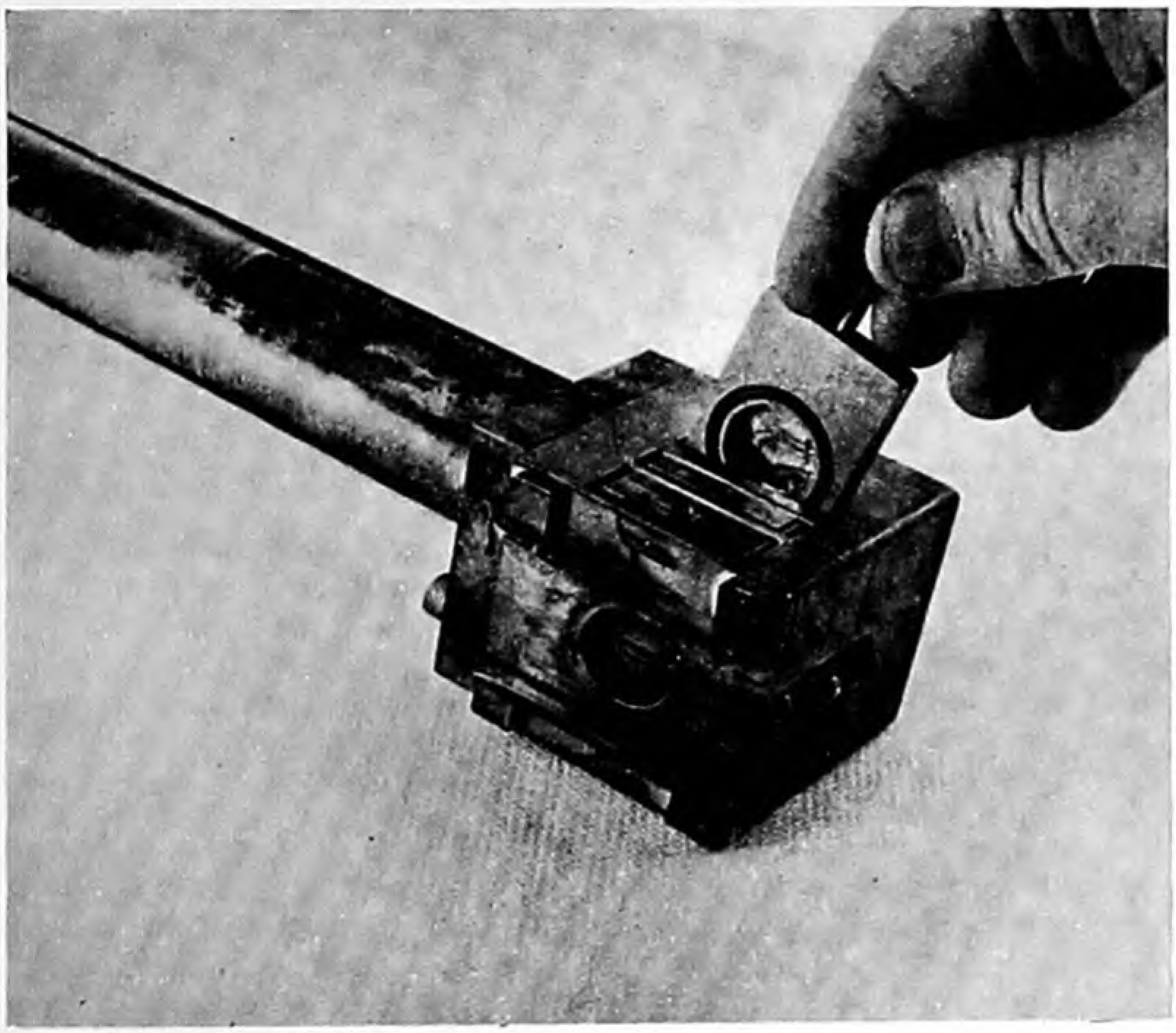
Fortunately there is a type of nuclear



IONIZATION CHAMBER measures the energy of particles emitted by radioactive atoms. It does so by emitting a pulse of electric current, the strength of which is proportional to the energy of the particle. In the photograph at the top a disk of the kind shown on page 360 is inserted into the ionization chamber. The drop of solution on the surface of the disk has been dried. At the right in the photograph below is the instrument which analyzes the pulses emitted by the chamber. The chamber is at left in this photograph.



reaction that does seem to offer hope for the production of heavier elements. This is the method of bombardment with projectiles heavier than alpha particles, which are helium nuclei. For example, isotopes of californium, einsteinium and fermium have already been produced by bombardment of uranium with nuclei of carbon, nitrogen and oxygen, respectively. These nuclei can be accelerated in cyclotrons of the conventional type, and the University of California and Yale University are building linear accelerators which will be devoted exclusively to the acceleration of heavy ions to energies sufficient to allow them to transmute the heaviest elements. The two machines are designed to produce rather substantial beams of all the nuclei up to neon, and possibly will be able to give usable beams of nuclei as heavy as argon.



CYCLOTRON TARGET of special design was used to bombard einsteinium in such a way that the mendelevium produced by the bombardment would be separated from it. The einsteinium was plated on gold foil; when the foil was bombarded in the cyclotron a few atoms of einsteinium were transmuted to mendelevium. The energy of the transmuting particles was sufficient to knock the atoms of mendelevium through the foil and deposit them on a second foil. Here the second foil is removed. The first foil is within the assembly.

The prediction of the chemical properties of the undiscovered elements beyond mendelevium seems to be quite straightforward. Elements 102 and 103, corresponding to ytterbium and lutetium in the rare-earth series, should complete the actinide group. This will involve filling what is known as the 5f shell of electrons, which, like the last shell (4f) of the rare-earth group, has 14 places. It is expected that element 104 will begin a new series, whose members will correspond to hafnium, tantalum, tungsten and so on, as the periodic table shown indicates [*see page* 359]. This series would end with the filling of the 6d electronic shell. The next series, assuming that heavier elements could be found (which, as we have noted, is very doubtful), would have a 7p shell and would close with hypothetical element 118. The chemical properties of all of these elements can be estimated from their postulated positions in the periodic table.

Thus the current chapter in the story of the synthetic elements has not yet come to its end. It is the hope of the authors that this article will already be out of date by the time many of our readers see it, because of the discovery of elements beyond mendelevium. The exciting field of modern alchemy is proceeding at such a pace that this possibility does not seem at all unlikely.

The Authors

ALBERT GHIORSO and GLENN T. SEABORG are leading workers in the University of California's Radiation Laboratory at Berkeley. One or both of them have participated in the discovery of all of the transuranium elements. Seaborg, a Nobel prize winner, is director of chemical research at the Radiation Laboratory. He received his Ph.D. from the University of California in 1937 and has been there ever since except for his work at the Metallurgical Laboratory of the University of Chicago during World War II: he was primarily responsible for the development of the chemical separation procedures used in the manufacture of plutonium. Besides the Nobel prize, he has received many honors: on January 11 he will be awarded the Perkin medal in industrial chemistry for 1957 at a dinner in his honor at the Waldorf-Astoria Hotel in New York. Ghiorso, an electrical engineer, has made major contributions in the instrumentation of nuclear chemistry and physics.

Bibliography

THE ACTINIDE ELEMENTS. Edited by Glenn T. Seaborg and Joseph J. Katz in *National Nuclear Energy Series*, Div. 4, Vol. 14A; 1954.

THE ACTINIDE SERIES. G. T. Seaborg in *Comprehensive Inorganic Chemistry*, Vol. 1, Chap. 3, pages 161-223; 1953.

SOURCEBOOK ON ATOMIC ENERGY. Samuel Glasstone. D. Van Nostrand Company, Inc., 1950.

THE TRANSURANIUM ELEMENTS. Edited by Glenn T. Seaborg, Joseph J. Katz and Winston M. Manning in *National Nuclear Energy Series*, Div. 4, Vol. 14B; 1949.

SCIENTIFIC AMERICAN June 1956, Vol. 194, No. 6, pp. 37-41 OFFPRINT 244

THE ANTI-PROTON

by Emilio Segrè and Clyde C. Wiegand

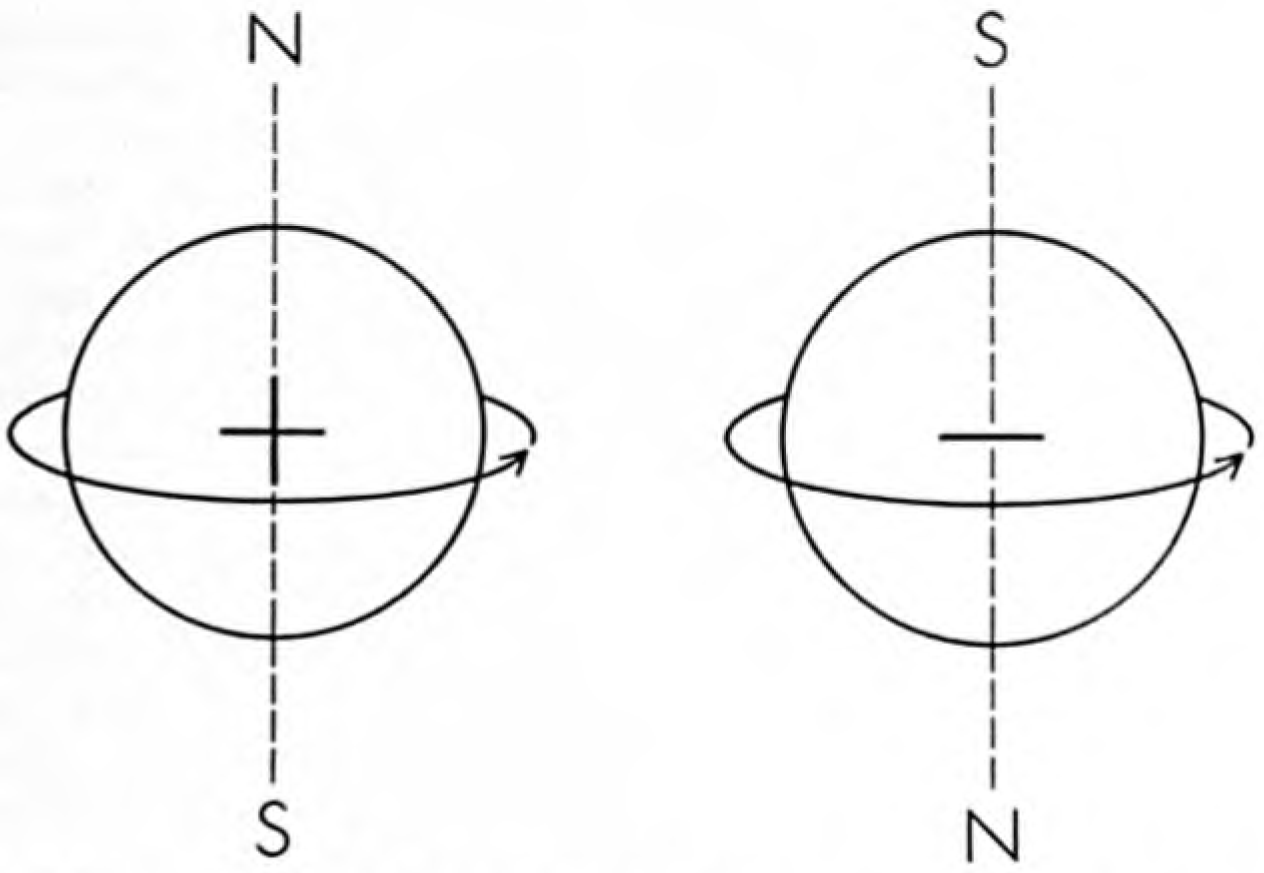
A quarter-century ago physical theory pointed to the existence of a fundamental particle of matter with the mass of a proton but the opposite charge. An account of its discovery in 1955.

In the past 10 years physicists have discovered many "elementary" particles (though we must frankly admit that we do not really know how to define an "elementary particle"). Not all of these discoveries have come unheralded. One might suppose that the only way to discover a particle should be by experiment, but this is not so, although of course experiment is the judge of last resort. Sometimes theoretical physicists, from hypothetical equations and calculations with pencil and paper, have predicted the existence of particles that had never been seen. These predictions, however strange some of them may seem, arise from a necessity to preserve basic principles which form the foundation of our present understanding of the physical universe. When necessary, physicists have been willing to entertain the existence of something never seen rather than to imperil these firmly established foundations. This article is the story of how such a prediction was verified.

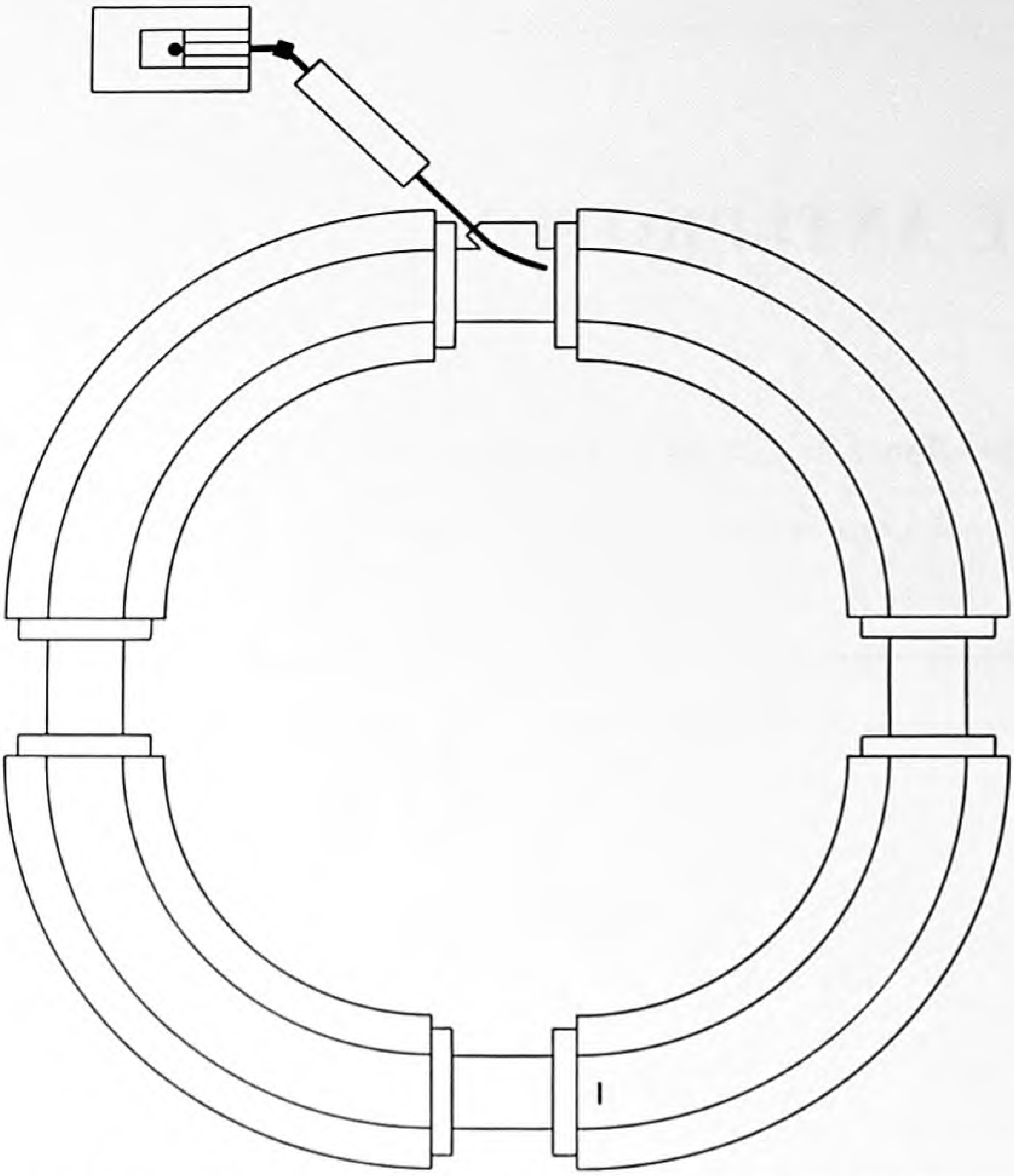
A quarter of a century ago P. A. M. Dirac of the University of Cambridge developed an equation, based on the most general principles of relativity and quantum mechanics, which described in a quantitative way various properties of the electron. He had to put in only the charge and mass of the electron—and then its spin, its associated magnetic moment and its behavior in the hydrogen atom followed with mathematical necessity. The fact that all this could be obtained automatically from one equation without *ad hoc* assumptions for each property was such a spectacular success that great faith was put in Dirac's equation and the theory on which it was based. Its discoverer found, however, that the equation required the existence of both positive and negative electrons: that is, it described not only the known negative electron but also an exactly symmetrical particle which was identical with the electron in every way except that its charge was positive instead of negative. It proved impossible to prevent Dirac's theory from giving both types of solutions. This meant that either Dirac's theory was wrong or there must be a positive electron which no physicist had ever detected or even suspected up to that time.

A few years after Dirac's prediction, Carl D. Anderson of the California Institute of Technology found positive electrons (positrons) among the particles produced by cosmic rays in a cloud chamber. This discovery not only was a triumph for Dirac's theory but also set physicists off on a new and more formidable search for another hypothetical particle—a search which was to take some 25 years and which was finally rewarded only a few months ago.

Dirac's general equation, slightly modified, should be applicable to the proton as well as to the electron. In this instance too it predicts the existence of an anti-particle—an antiproton identical to the proton but with a negative instead of a positive charge. The unknown particle's symmetry to the proton clearly defines some of its properties. A particle



PROTON (*left*) may be regarded as a spinning sphere weighing 1.6724 x 10^{-24} grams. It has positive electric charge and, as a consequence of its rotation, north and south magnetic poles. The antiproton (*right*) has the same spin and mass. It has the same amount of electric charge, but of the opposite sign. Its north and south magnetic poles are similarly reversed.



BEVATRON at the University of California produced antiprotons by accelerating protons to 6.2 billion electron volts. This schematic plan view shows the four magnet-enclosed segments in which the protons are accelerated. The radius of each segment is 50 feet. The protons are injected into the machine by two accelerators at the top. The copper target in which the antiprotons are produced is represented by the heavy vertical line at the bottom right.

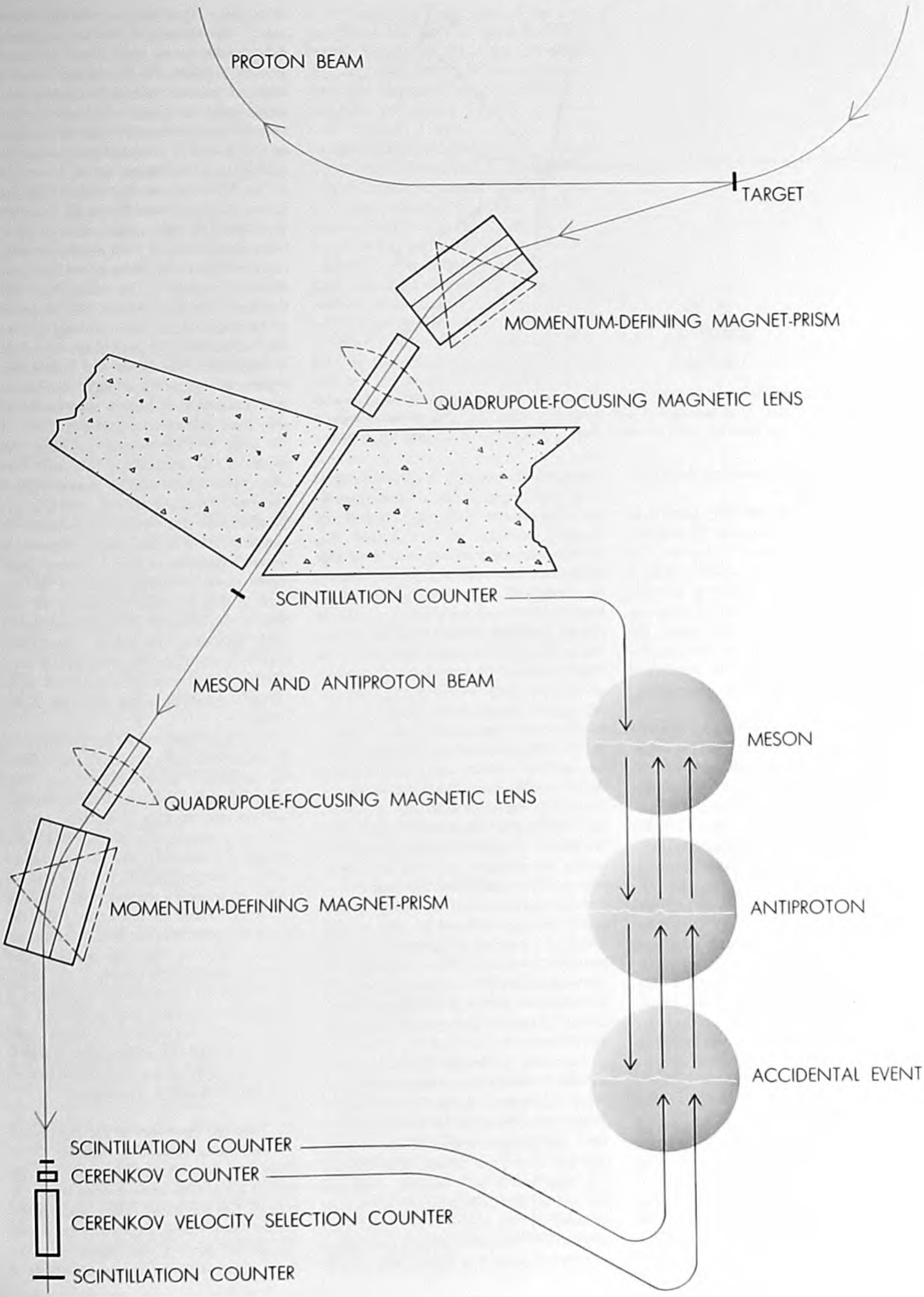
in order to have the right to be called an antiproton: (1) must have the same mass as a proton (1.6724×10^{-24} grams); (2) must have an equal charge of opposite sign (4.8028×10^{-10} electrostatic units); (3) must be stable, in the sense that it will not decay spontaneously into a different particle and will last forever in a vacuum; (4) must disappear in a mutual annihilation when it encounters a proton or a neutron, liberating energy equivalent to the mass of the two particles; (5) is never generated separately but only in a pair with a proton or neutron; and (6) must have an angular momentum (spin) equal to that of the proton. Like the proton, an antiproton must also have a magnetic moment (*i.e.*, behave like a little magnet), and when it spins in the same direction as a proton its magnetic moment is equal in magnitude but of opposite sign to that of the proton; that is, the "north and south" poles are reversed.

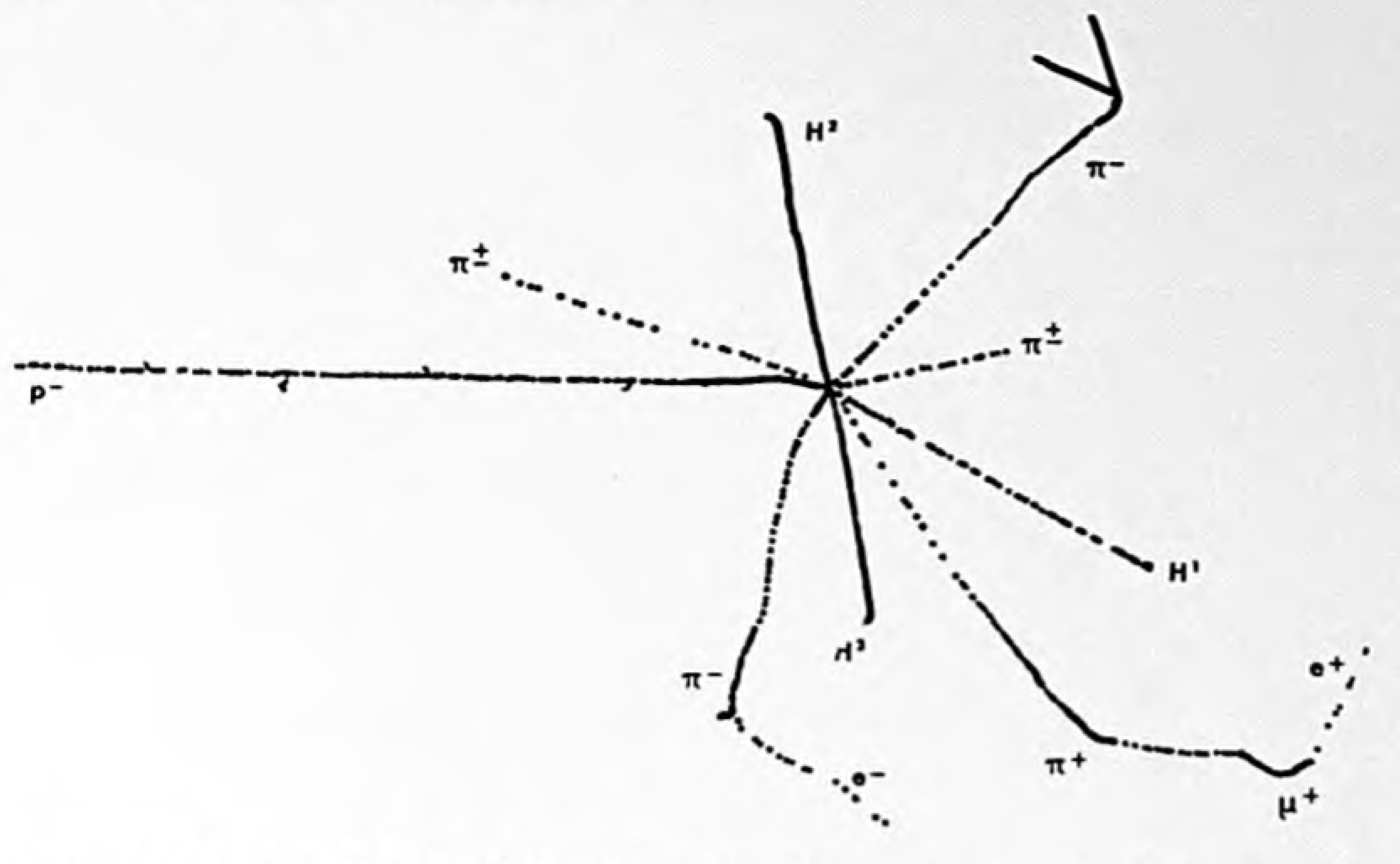
With all these clues, physicists naturally began an intensive search for the antiproton. Since it was apparent that creation of the particle required tremendous energy, the most likely place to look for it was in cosmic rays. On a few occasions investigators found events which seemed to signal the generation of an antiproton, but there was never sufficient information to identify it with certainty. The question then arose as to how much energy would be needed to create antiprotons in the laboratory with an accelerator.

Because an antiproton can be created only in a pair with a proton, we need at least the energy equivalent to the mass of two protons. Albert Einstein's theorem, $E=mc^2$, tells us that this amounts to 2×938, or 1,876 million electron volts (*i.e.*, about two billion electron volts). However, we need much more than two Bev in the proposed laboratory experiment. To convert energy into particles we must concentrate the energy at a point; this is best accomplished by hurling a high-energy particle at a target—*e.g.*, a proton against a proton. After the collision we shall have four particles: the two original protons plus the newly created proton-antiproton pair. Each of the four will emerge from the collision with a kinetic energy amounting to about one Bev. Thus the generation of an antiproton by this method takes two Bev (creation of the proton-antiproton pair) plus four Bev (the kinetic energy of the four emerging particles). It was with these numbers in mind that the Bevatron at the University of California was designed. It was built to accelerate protons to a kinetic energy of more than six Bev, with the hope of producing antiprotons.

When the Bevatron began to bombard a target made of copper with six-Bev protons, the next problem was to detect and identify any antiprotons created. A plan for the search was devised by Owen Chamberlain, Thomas Ypsilantis and the authors of this article. The plan was based on three properties which could conveniently be determined. First, the stability of the particle meant that it should live long enough to pass through a long apparatus. Second, its negative charge could be identified by the direction of deflection of the particle by an applied magnetic field, and the magnitude of its charge could be gauged by the amount of ionization it produced along its path. Third, its mass could be calculated from the curve of its

EXPERIMENTAL SETUP which demonstrated the existence of the antiproton is depicted on opposite page. The colored line at the top is the orbit of the protons in the Bevatron. The path of the antiprotons and other particles produced in the target is traced by the colored line from upper right to lower left. From the target to the last scintillation counter the particles travel 80 feet. At the lower right the various events of the experiment are represented by their characteristic cathode ray traces. Above the center of the diagram is the concrete shield of the Bevatron. The prisms and lenses superimposed with broken lines on the four magnets symbolize their function. The experiment is described in detail by the text of this article.


PROTON BEAM
TARGET
MOMENTUM-DEFINING MAGNET-PRISM
QUADRUPOLE-FOCUSING MAGNETIC LENS
SCINTILLATION COUNTER
MESON AND ANTIPROTON BEAM
MESON
QUADRUPOLE-FOCUSING MAGNETIC LENS
ANTIPROTON
MOMENTUM-DEFINING MAGNET-PRISM
ACCIDENTAL EVENT
SCINTILLATION COUNTER
CERENKOV COUNTER
CERENKOV VELOCITY SELECTION COUNTER
SCINTILLATION COUNTER



NUCLEAR EMULSION PHOTOGRAPH reproduced by tracing shows the death of an antiproton (p^-) in a "star" of pi mesons (π) and hydrogen nuclei (H). One of the pi mesons has decayed into an electron (e); another into a mu meson (μ) and an electron.

trajectory in a given magnetic field if its velocity was known.

The trajectory of a charged particle in a magnetic field depends on its momentum: once the trajectory is known, the momentum can be calculated. Now if we also measure the velocity of the particle (say by timing its travel between two given points in the apparatus), we can compute the mass from the momentum and velocity, using the relativistic equation which connects momentum, rest mass and velocity.

All this sounds very simple, but the main difficulty in the experiment arises from a complication which we have thus far neglected to mention. When the beam of 6.2-Bev protons hits the target, it generates a great many other particles which have the same momentum as the antiprotons. Most of them are mesons, the particles supposed to represent the cement that holds the nucleons together in the nucleus. It turned out that there were about 40,000 mesons for each rare antiproton in the stream of emerging particles focused by our magnets. The mesons follow exactly the same trajectory as the antiprotons, but they are lighter and travel with a velocity practically identical to that of light, whereas the heavier antiproton moves with 78 per cent of the velocity of light. The problem was to pick out of the stream the occasional heavy particle (one in 40,000) moving with the right velocity to be an antiproton.

An extensive array of bending magnets, magnetic focusing lenses and detectors was set up to comb out antiprotons [*see diagram on page 371*]. From the spray of charged particles emerging from the copper target a bending magnet first sorted out the negatively charged particles of the desired momentum, bending them in a particular trajectory. This stream was then focused by a magnetic lens. The focused beam now encountered a detector—a disk of plastic material which scintillates when charged particles pass through. The main purpose of the detector was to serve as a "stop watch" for timing the passage of particles so as to measure their velocity: precisely 40 feet farther on they hit a second scintillating detector, and the velocity was reckoned from the time taken to travel the distance between the two "stop watches." The flashes of light from each scintillator were translated by photosensitive tubes into pulses of electric current, and these pulses were recorded as pips on a cathode ray screen. This timing system could measure differences of one billionth of a second in the travel time of particles over the 40-foot interval. In our experiment the antiprotons cross the 40-foot distance in 51 billionths of a second, whereas the mesons take only 40 billionths of a second.

However, we found that we needed an independent measurement of the particles' velocities as a check against accidental coincidences. So many mesons were streaming through our "speed trap" that sometimes one meson triggered the first stop watch and another triggered the second after an interval that corresponded to the travel time of an antiproton. We therefore placed a velocity-selecting counter just beyond the second scintillator. This unique selector makes use of the Cerenkov effect, discovered many years ago by the Russian physicist Pavel Cerenkov. He found that when a charged particle passes through a medium such as glass or quartz with a velocity greater than the velocity of light in that medium, it emits light—an effect analogous to the shock waves produced in air by a jet aircraft exceeding the speed of sound. Now the angle at which the Cerenkov light radiation is emitted, with respect to the path of the particle, depends upon the velocity of the particle. An analogy is the wake of a boat: the faster the boat travels, the narrower is the angle of its wake. Taking advantage of this fact, we put a piece of quartz in the path of the beam and arranged a system of mirrors and light shields so that Cerenkov radiation was recorded only from particles traveling at 75 to 78 per cent of the velocity of light—the speed of the antiproton. We took two other precautions against spurious identification. To make sure of weeding out mesons and other unwanted particles we placed in front of the velocity selector a "guard" detector of the Cerenkov type which gave a warning signal of the arrival of any particle exceeding 78 per cent of the velocity of light, and to exclude particles that might come from outside the system we used a final scintillation counter which recorded only particles traveling in the direction of the beam.

Thus a particle would be identified as an antiproton only when all the following conditions were fulfilled: the "stop watch" counters indicated that a particle had passed through with the correct velocity (crossing the 40 feet in 51 billionths of a second); the guard counter gave no warning signal; the velocity selector registered a particle with velocity between 75 and 78 per cent of the velocity of light; and the final scintillation counter showed that the particle had coursed through the length of the selector. When all these things happened, a characteristic sweep was traced on the oscilloscope [*see page 371*]. Many more tests were made to confirm that this type of sweep really meant that an antiproton had passed through the system.

When the discovery of the antiproton was announced last October, 60 of them had been recorded, at an average rate of about four to each hour of operation of the Bevatron. They had passed all the tests which we had preordained before the start of the experiment. We were quite gratified by the comment of

a highly esteemed colleague who was visiting from another university where he had just finished an important and difficult experiment on mesons. After examining our tests, he said, "I wish that my own experiments on mu mesons were as convincing as this." At this time several long-standing bets on the existence of the antiproton started to be paid. The largest we know of was for $500. (We were not personally involved.)

It was still highly desirable to have some information on the process of annihilation of the antiproton when it encountered a proton. The first experiment along this line was performed by a group consisting of J. Brabant, B. Cork, N. Horowitz, B. Moyer, J. Murray, R. Wallace and W. Wenzel. They arranged to trap an antiproton from our apparatus in a piece of glass. On being stopped, the antiproton, and the proton which presumably was annihilated with it, emitted charged particles which moved fast enough to release considerable light by Cerenkov radiation. A study of this light confirmed that the particle selected as an antiproton was definitely not a meson.

While all this was going on, another experiment for detection of the antiproton was started by the authors, Chamberlain, W. Chupp and G. Goldhaber, in collaboration with a team of physicists in Italy: E. Amaldi (a former fellow student with Segrè of the late Enrico Fermi), G. Baroni, C. Castagnoli, C. Franzinetti and A. Manfredini. It was decided to try to find the tracks of antiprotons in photographic emulsions. If we could detect them there, we could get direct information about the antiproton's destruction.

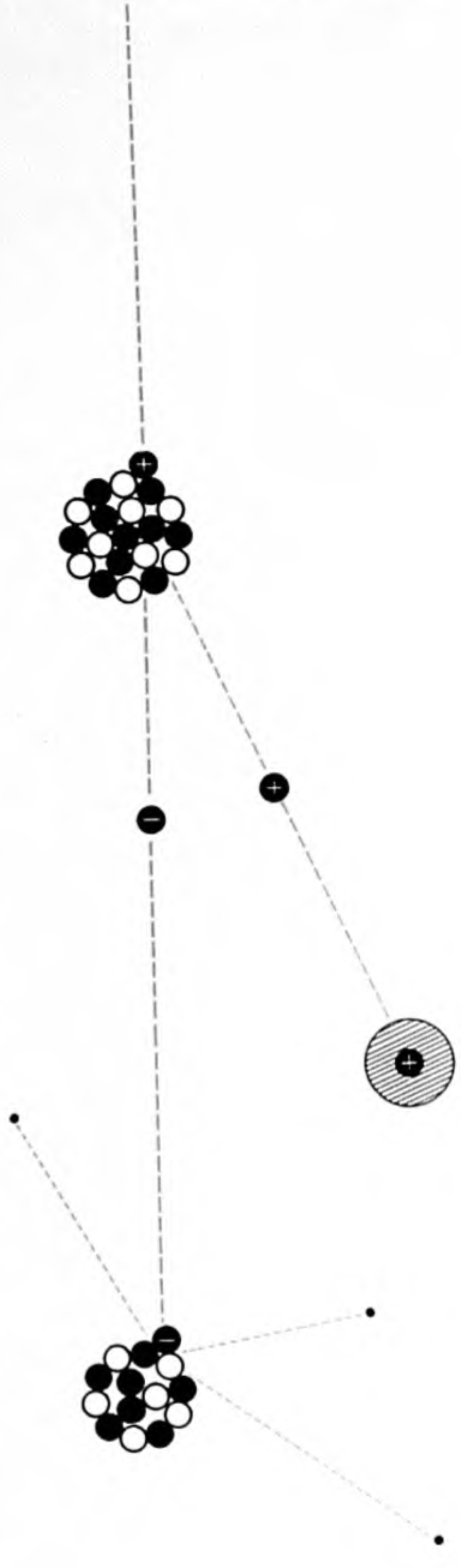
We exposed photographic plates in a beam which should yield antiprotons and then sent some of the plates to Rome and examined some ourselves in Berkeley. In spite of strenuous efforts by both groups, only one star that might represent a proton-antiproton annihilation was found—by the scanners in Rome. Later experiments by our group (including D. Keller and H. Steiner) indicated that absorbers which we had used to slow down the antiprotons before they entered the photographic plates had unexpectedly destroyed many of the antiprotons. The absorbers were then removed and new plates were exposed, with the result that tracks of about 20 antiprotons have now been detected in emulsions by observers in Berkeley.

A star of annihilation is pictured at the top of the opposite page. The track of the incoming antiproton has the range predicted for it. The particle lost its kinetic energy by collisions and ionization and came to rest in the emulsion. It was then captured by a nucleus in the emulsion and promptly annihilated itself with another nucleon. Many of the charged fragments into which it broke down can be identified by their tracks; others cannot be named with certainty; still others were neutral particles which made no tracks in the plate. At all events, we know for certain that the total energy released in the annihilation was greater than the mass equivalent of the antiproton—proof that another nucleon was annihilated along with it.

As usual with all discoveries, the advent of the antiproton has launched a host of new questions, on which work is progressing. For the time being only Berkeley has an accelerator powerful enough to produce antiprotons. The next to enter the picture, according to reports at the Geneva conference of last summer, should be a U.S.S.R. machine, now in an advanced stage of construction.

An interesting subject for contemplation is the possible existence of an "anti-world." This would be a world in which all particles are opposite in charge to our own: the hydrogen atom, for instance, would have an antiproton as its nucleus and a positron in place of the electron. We know of no method by which we could recognize the existence of such a universe by astronomical observation. But if antimatter exists and if it should come into contact somewhere with ordinary matter as we know it, the two forms of matter would annihilate each other with a huge release of energy, mostly as mesons. Whether we would see this event would depend on the density of the matter colliding. If it were spread out as thinly as the average density of matter in the galaxies, the effect might not be very conspicuous. It is also possible that a collision even between concentrated masses of matter and antimatter would not be very spectacular astronomically, for they probably would repel each other, by radiation pressure, as soon as they came into contact.

If the universe originated from the transformation of pure energy into nucleons and electrons, we must suppose, in order to preserve the principle of the constancy of the number of these particles, that somewhere there are antinucleons and antielectrons equal in number to those of our world. It is a speculation which gives a highly satisfying symmetry to creation.



LIFE CYCLE of an antiproton is schematically depicted. A high-energy proton (*black ball with plus sign at the top*) collides with another proton in the target nucleus. This gives rise to a new proton (*black ball with plus sign at right center*) and an antiproton (*black ball with minus sign*). The new proton travels until it comes to rest as the nucleus of a hydrogen atom (*right*). The antiproton continues until it encounters another nucleus (*bottom*). The antiproton and a proton or a neutron are then both annihilated in a shower of various particles.

The Authors

EMILIO SEGRÈ and CLYDE E. WIEGAND both nuclear physicists at the University of California, are members of the team that discovered the antiproton. Segrè, born in Italy in 1905, took his doctor's degree in physics at the University of Rome under Enrico Fermi in 1928. He worked with Fermi on the radioactivity of neutrons from 1934 to 1936. From there he went to Palermo as director of the physics laboratory. While at Palermo he and C. Perrier announced the discovery of technetium, the first artificially made element. He has been at the University of California in Berkeley since 1938, except for three years at the Los Alamos Scientific Laboratory during World War II. Segrè seems to have a divining rod for discovering elements and particles. Besides the antiproton and technetium, he identified the first plutonium with Glenn T. Seaborg, J. W. Kennedy and A. C. Wahl, and found the artificial element astatine with Dale Corson and K. R. MacKenzie. Wiegand is a physicist at the Radiation Laboratory of the University of California. Born in the State of Washington, he graduated from Willamette University in Oregon in 1940. He became a graduate student of Segrè at California, went with his professor to Los Alamos, and took his Ph.D. in 1950 after their return to Berkeley.

Bibliography

ANTIPROTONS. O. Chamberlain *et al.* in *Nature*. Vol. 177, No. 4497, pages 11-12; January 7, 1956.

OBSERVATION OF ANTIPROTONS. Owen Chamberlain *et al.* in *The Physical Review*, Vol. 100, No. 3, pages 947-950; November 1, 1955.

SCINTILLATION COUNTERS. George B. Collins in *Scientific American*, Vol. 189, No. 5, pages 36-41; November, 1953.

SCIENTIFIC AMERICAN August 1956, Vol. 195, No. 2, pp. 76-84 OFFPRINT 245

THE EFFECTS OF RADIATION ON SOLIDS

by Frederick Seitz and Eugene P. Wigner

The orderly atomic arrangement characteristic of metals and other crystals determines many of their properties. Energetic radiation disturbs the order and thus can drastically alter the properties.

In the anxious days when atomic scientists were building the first chain-reacting pile in the "Metallurgical Laboratory" at the University of Chicago, no problem was more worrisome than the question concerning how the pile would be affected by its own radiation after it became active. On most of the other problems—the critical requirements for the chain reaction, controls, shielding, cooling—the physicists felt fairly confident of their calculations. But the radiation question was full of uncertainties. It was known that exposure even to weak natural radioactivity could change the structure and properties of materials. What would happen to the uranium rods in the reactor under the disruptive forces of intense neutron radiation, nuclear fissions and so on? More serious still, what would happen to the graphite moderator? Graphite was a part of the actual structure of the pile; unlike the uranium, it was not to be removed or replaced from time to time; and it was known to be subject to damage by radiation.

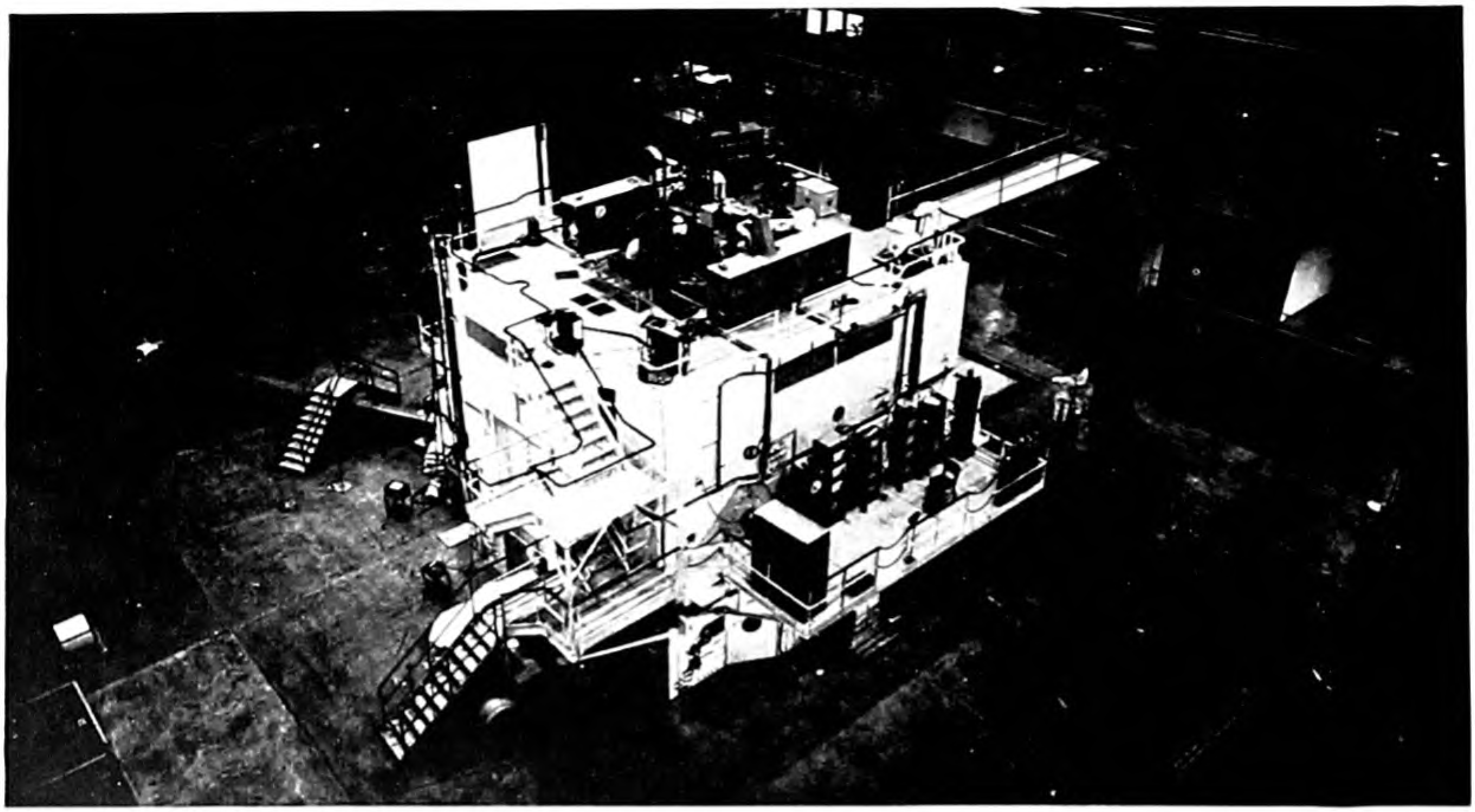
The group concerned with the future health of the new atomic "child" was so uncertain and pessimistic about the reactor's ability to survive radiation and other "diseases" that it reported: "It would be unscientific to claim a useful life longer than about 100 days." More than 50 times that period has now passed and nearly all the original reactors are still alive and operating. What we did not realize at the time was that graphite, as well as metal, has some ability to recover from radiation damage—to heal its wounds, so to speak. Nevertheless, the effect of radiation on solids remains an important and absorbing study. It is still a major practical problem in the construction of reactors; besides this, it has become a valuable tool for fundamental research into the properties of solids. Research on radiation damage is now being carried on not only in the national laboratories of the Atomic Energy Commission but also at a number of universities and industrial laboratories. The AEC recently announced eight such research contracts totaling well over $250,000 a year. The program of study of radiation effects on solids has steadily grown both in magnitude and in scope.

Let us try to describe some of the facts we have learned about radiation damage. Metals and nonmetals react differently; we shall consider first the effects on a nonmetal—the graphite (crystalline carbon) commonly used as the moderator in a reactor. The neutrons released by uranium fission in a reactor have a kinetic energy of about one million electron volts. When a fast neutron strikes the nucleus of a carbon atom in the moderator, it transfers a substantial fraction of its kinetic energy to the atom, and the latter recoils from the impact. Since the carbon atom's recoil energy is much greater than the binding energy holding it in the crystal lattice (which is less than 10 electron volts), the atom is thrown out of its normal position. This results in two defects in the lattice: the dislodged atom occupies an interstitial space in the lattice (like a marcher out of his row in a parade), and it leaves behind a vacant site in the regular order.

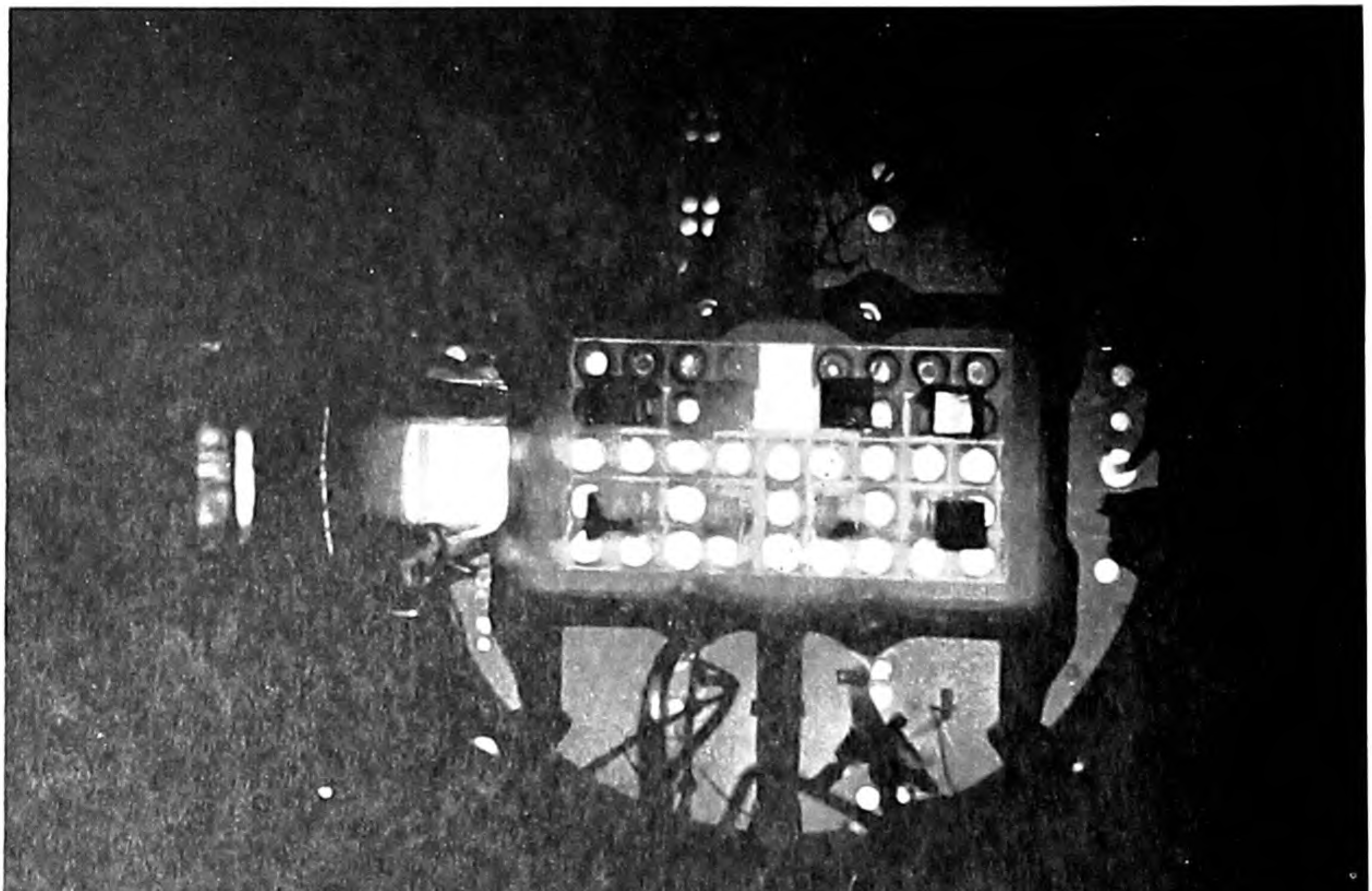
The foregoing describes the direct effect of collisions between fast neutrons and atoms in the lattice. These collisions in themselves account for only a small part of the damage actually produced. A fast neutron dislodges about 60 carbon atoms, at most, before it is slowed to a harmless speed. It is the recoiling carbon atoms that produce most of the damage in the lattice. They have bulk as well as speed. The first carbon atom hit by a million-volt neutron, for example, recoils with an energy of about 150,000 electron volts. In effect it acts like a strong and husky man who decides to get out of a very crowded subway rather suddenly. It throws the other atoms to right and left until it reaches the end of its range, that is, until its energy is exhausted.

Now it develops that in the atomic world this series of events takes a turn which is the opposite of what one might expect if he thinks in terms of mechanical collisions. The charging atom creates more havoc near the end of its rush than it does at the beginning. The reason is that we are dealing here with interatomic forces rather than what we usually think of as physical contact. As the fast-moving atom begins its dash through the crowd of surrounding atoms, its encounter with each one is too fleeting to permit much transfer of its momentum. It therefore dislodges only an occasional atom from its lattice position. But as the traveling atom slows down, the interatomic forces have more time to act, and it displaces more and more atoms. Finally, when it drops to a certain low velocity, it transfers its remaining energy to a local cluster of atoms. As a result the tiny local region suddenly heats up, sometimes to a temperature as high as 10,000 degrees centigrade. This phenomenon, called a "thermal spike" or "displacement spike," lasts only about one hundredth of a billionth of a second, but it may damage or deform the crystal.

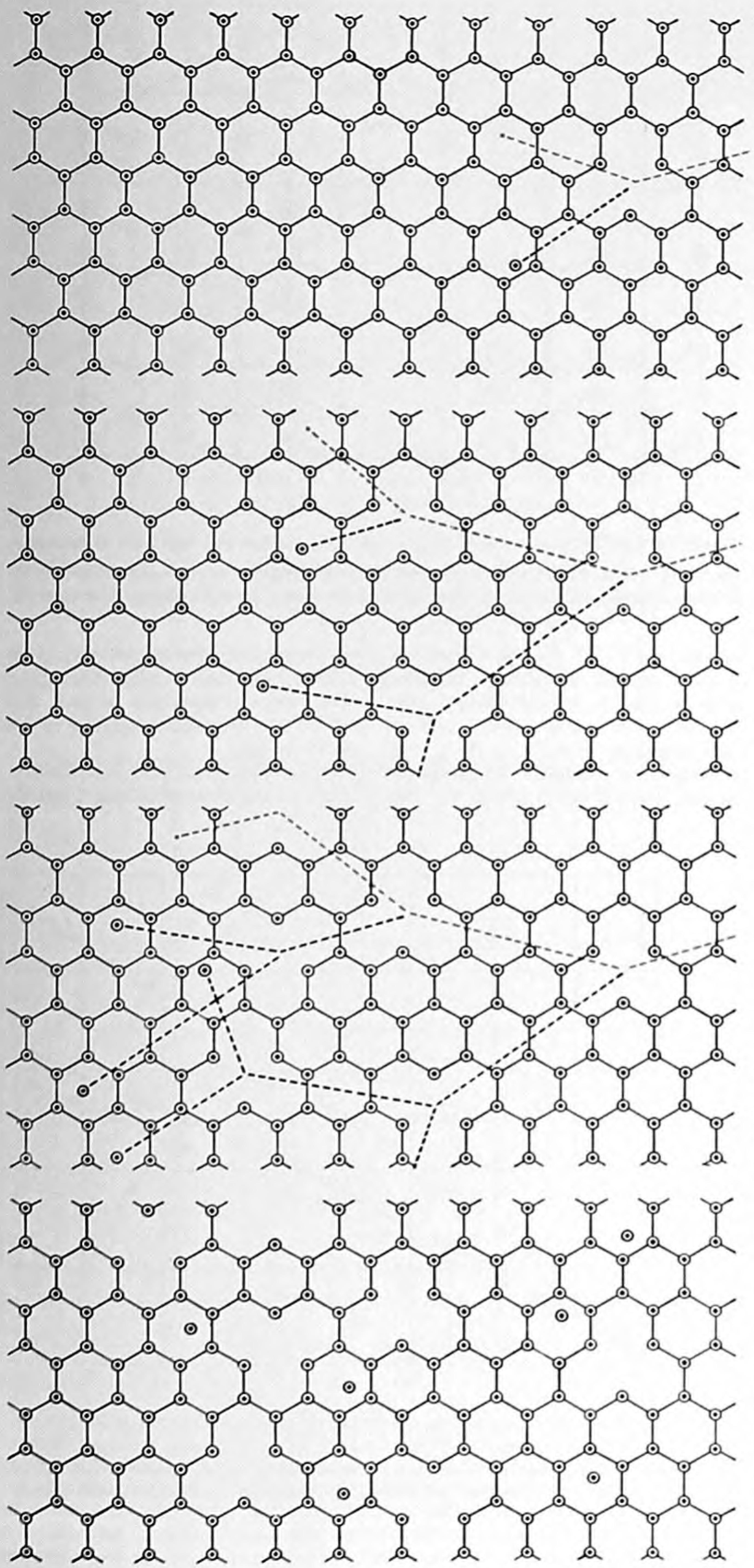
Its effects are fairly complicated and not yet well understood. It appears that the minute "spike" region melts. Evidence of this melting has been found in radiation experiments on a carefully prepared alloy of copper and zinc. The atoms were arranged in a regular lattice in which each copper atom was surrounded by eight zinc atoms and *vice*



MATERIALS TESTING REACTOR at the National Reactor Testing Station in Arco, Idaho, is used to study the effects of radiation on metals and other substances. **This research is leading to improved radiation-resistant materials for use in future reactors.**



IRRADIATED MATERIALS are photographed in the interior of the Materials Testing Reactor through 20 feet of water. Most of the glowing circles are fuel elements. **The light comes from Cerenkov radiation, produced in transparent materials by very fast particles.**



versa. Bombardment of this crystal with neutrons was done at very low temperatures, near the temperature of liquid helium, in order to "freeze in" any changes in the crystal. Analyses afterward showed that the atoms had become mixed in a disordered way, and that most of the disordering must have taken place in regions of thermal spikes.

Besides melting, the heated regions expand. Such swelling causes deformations of the crystal, some of which presumably remain after the hot regions cool, so that the material around them is permanently distorted.

In a crystal damaged by radiation it is very difficult to distinguish how much of the damage is due to these spikes and how much to simple displacement of atoms. We can assume that spikes are a more important source of damage in metals than in graphite, because in the heavier elements recoiling atoms produce spikes at a higher energy level and therefore have a larger fraction of their energy left for producing them. In the case of graphite, the moving carbon atoms have used up most of their energy dislodging atoms before they drop to the low velocity at which they generate spikes. We can estimate that the most damaging part of the flight of recoil atoms in graphite is in the velocity range from 100,000 down to 10,000 electron volts.

It has become clear that radiation can produce a great variety of defects in the lattice, resulting in varying damage to the material.

In this account of the process that produces radiation damage in solids we have given most attention to graphite, but much of what we have said applies to the metal fuel in a reactor as well. The agent of damage is essentially the same: namely, flying particles. The principal difference is that in uranium the important bombarding particles are not neutrons but fission products. The heavy fission fragments hit atoms in the crystal lattice far harder than neutrons, and the atoms receive, on the average, about

LATTICE DEFECTS are produced when a neutron strikes a graphite crystal. The hexagonal crystal structure is represented diagrammatically in two dimensions. At top the neutron (*colored dot*) has struck and dislodged a single atom. The next two drawings show how the process builds up, with both neutron and recoiling atoms acting to dislodge further atoms (*neutron path is in color; atom paths are in black*). At bottom is the final result: a lattice with a number of vacant sites and "interstitial" atoms.

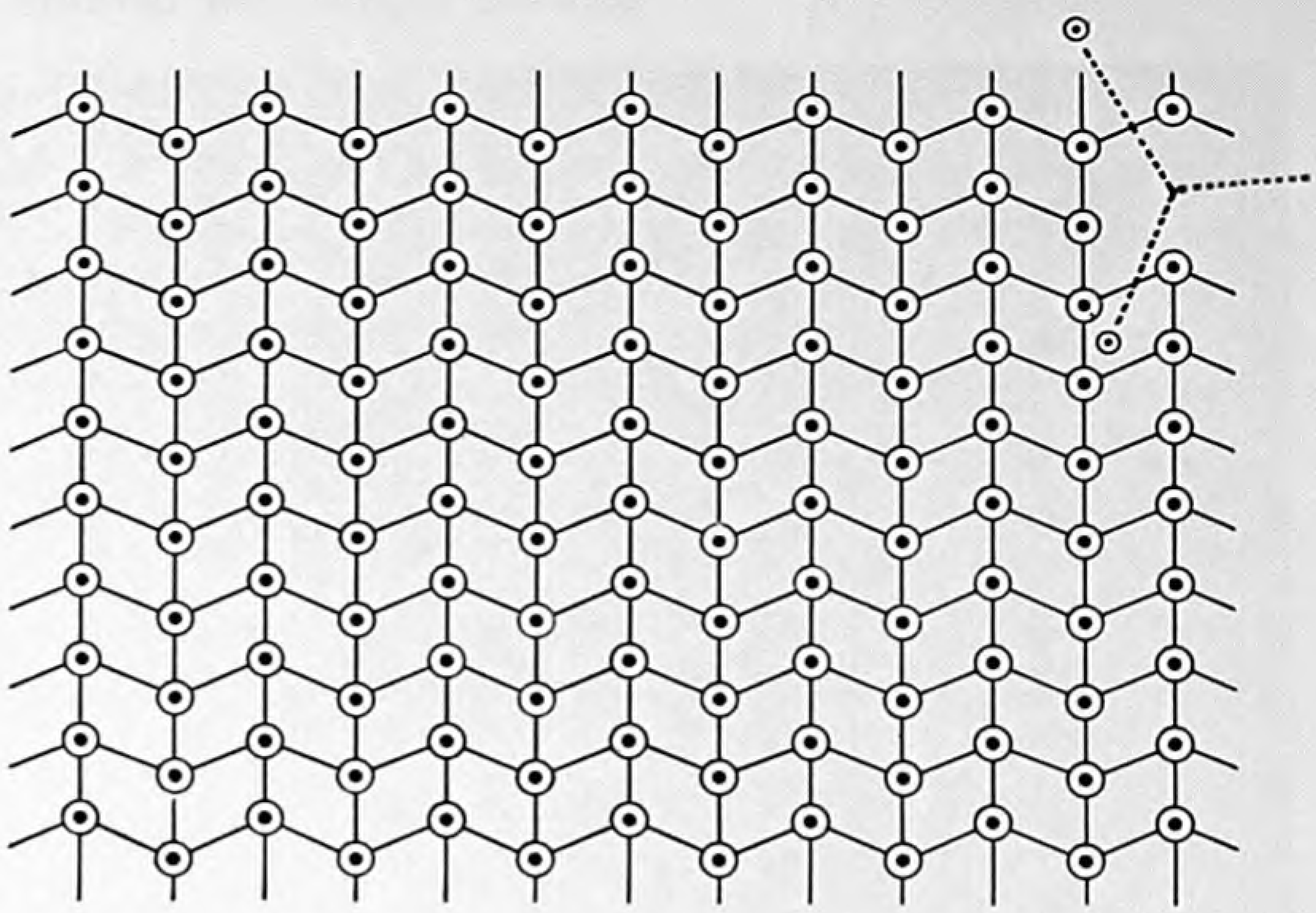
1,000 times more energy. The damage is therefore much greater. In addition, the fission conversion of part of the uranium into other elements also weakens the metal. Fortunately metals are tough and can stand a lot, particularly if they do not have to stand it too long!

Let us consider now the recuperating powers of materials damaged by radiation. Usually dislodged atoms attempt to return to something resembling their original positions in the lattice and to restore their original properties. We can investigate the recovery process best at low temperatures. If the damaged material is held at a temperature where the atoms can move around a bit, the interstitial atoms and vacancies will begin to recombine and the lattice distortion will heal. The crystal's properties then tend to return to normal. This is well illustrated by a study of the recovery of copper after it was irradiated near the temperature of liquid helium. The property measured was its conduction of electricity. Copper, which is a nearly ideal metal, recovers very rapidly if it is irradiated near room temperature; to "freeze in" all the damage and prevent recovery during irradiation it must be kept not far above absolute zero. Now when the temperature of the irradiated specimen is raised to about 35 degrees Kelvin, its electrical conductivity increases sharply. It is not yet known whether this abrupt and irreversible change, common to copper and many other metals, is a result of the reunion of vacancies and interstitial atoms which are very close to one another or whether it is due to healing of some of the distortion produced by thermal spikes. This is one of the critical questions being investigated at several laboratories.

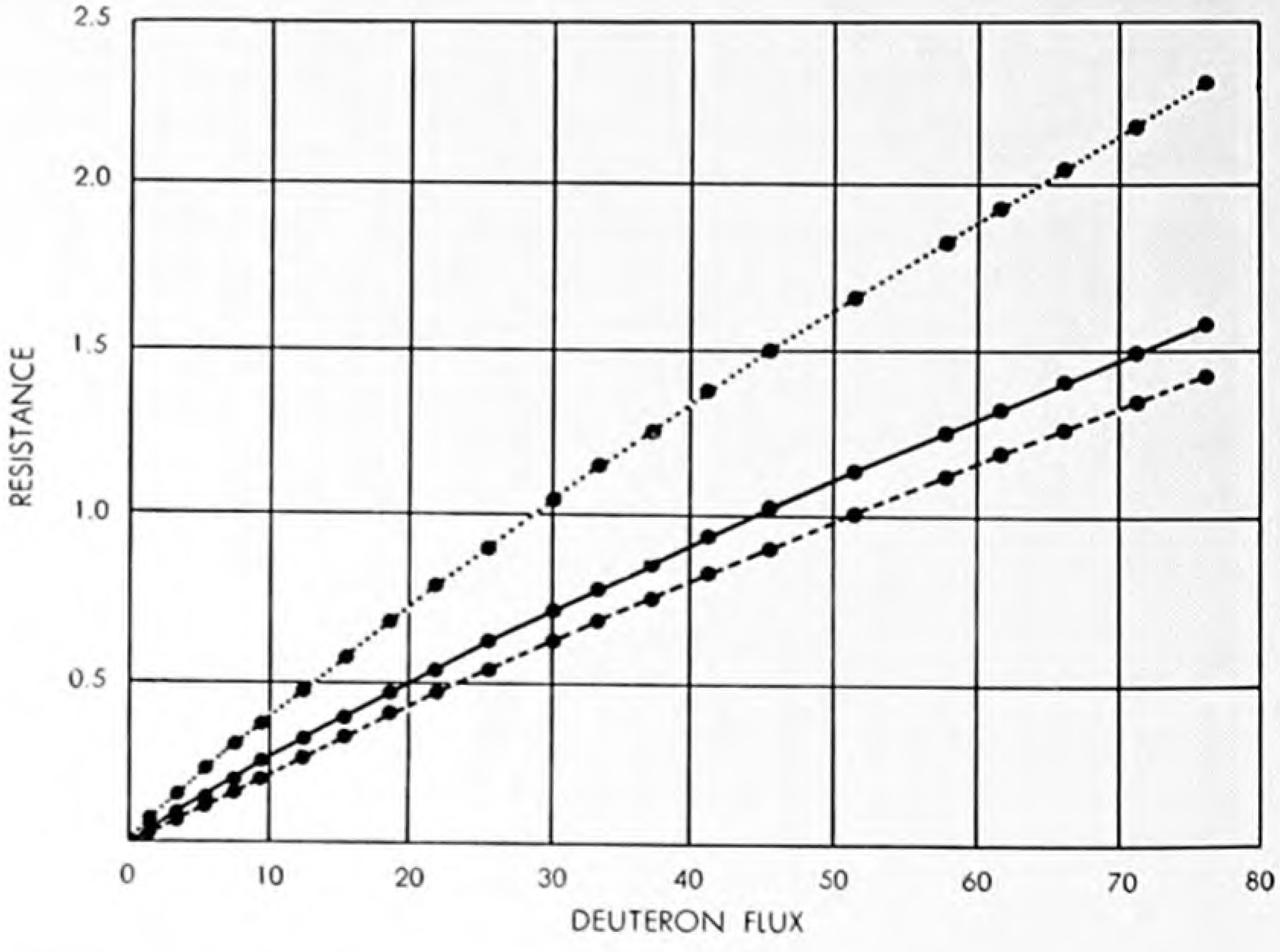
It is interesting to note that each increase in temperature permits a little more of the damage to heal. This shows that there is a spectrum of different types of defects, some of which are more resistant to correction than others. We know that small traces of impurity atoms can have a significant influence on the rate of recovery. Some of the defects produced are so stable that one must heat the metals to temperatures nearly halfway to the melting point in order to remove them.

On the whole, pure metals are the most resistant of all materials to radiation damage and recover most easily, presumably because the atoms in metals are most mobile. But reactions like those in metals have been found in valence compounds such as diamond, silicon and germanium, and in simple salts and oxides such as sodium chloride and beryllium oxide. On the other hand, organic materials, particularly polymers such as plastics, are exceedingly sensitive to radiation and suffer permanent and irreparable changes, In these cases the damage is associated with the breaking of chemical bonds which are difficult to rejoin in the original way. Most polymers lose their ductility when given even moderate exposures. In brief, they behave in a way almost opposite to that of the metals.

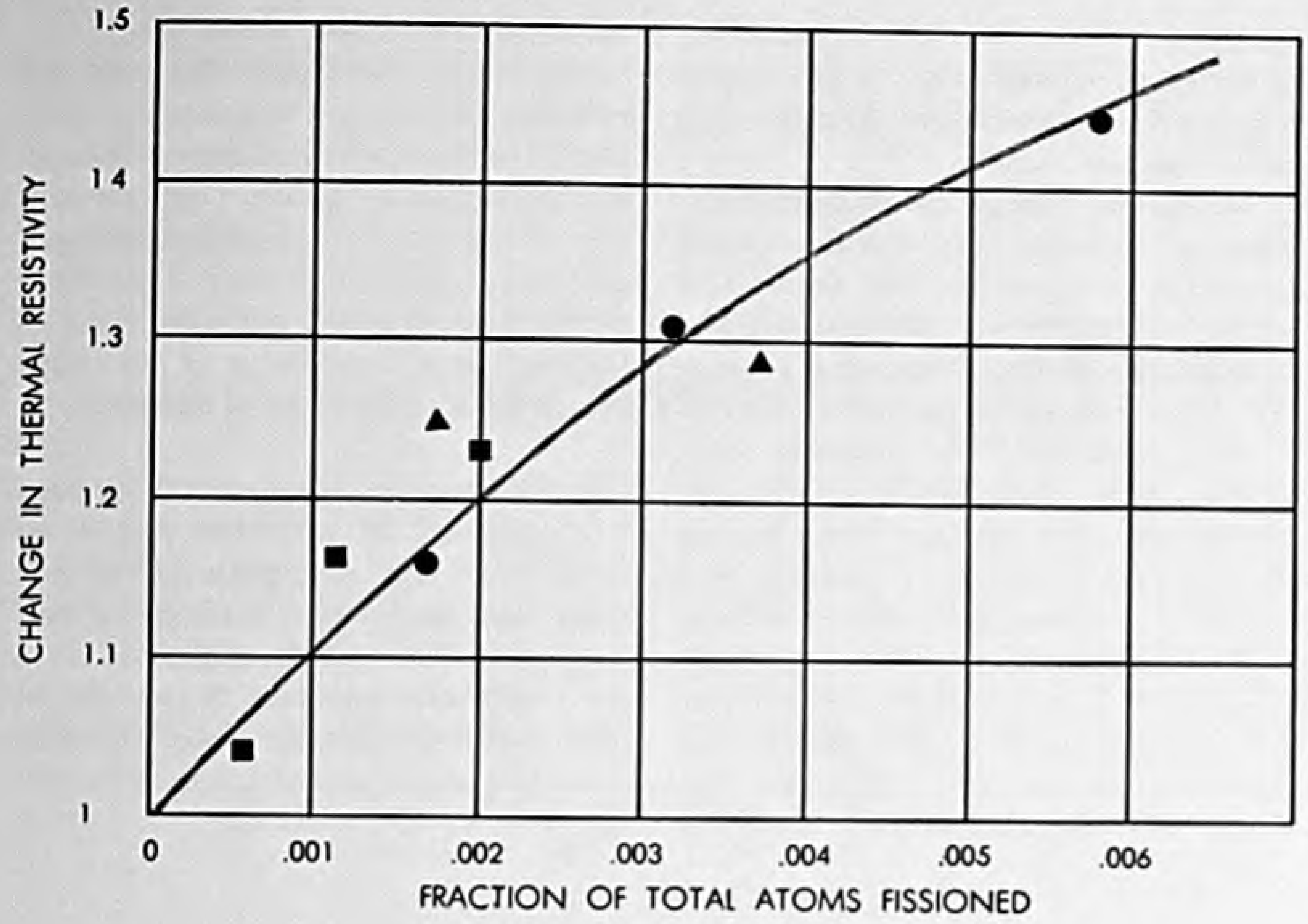
From the practical standpoint, what are the types of damage—and the

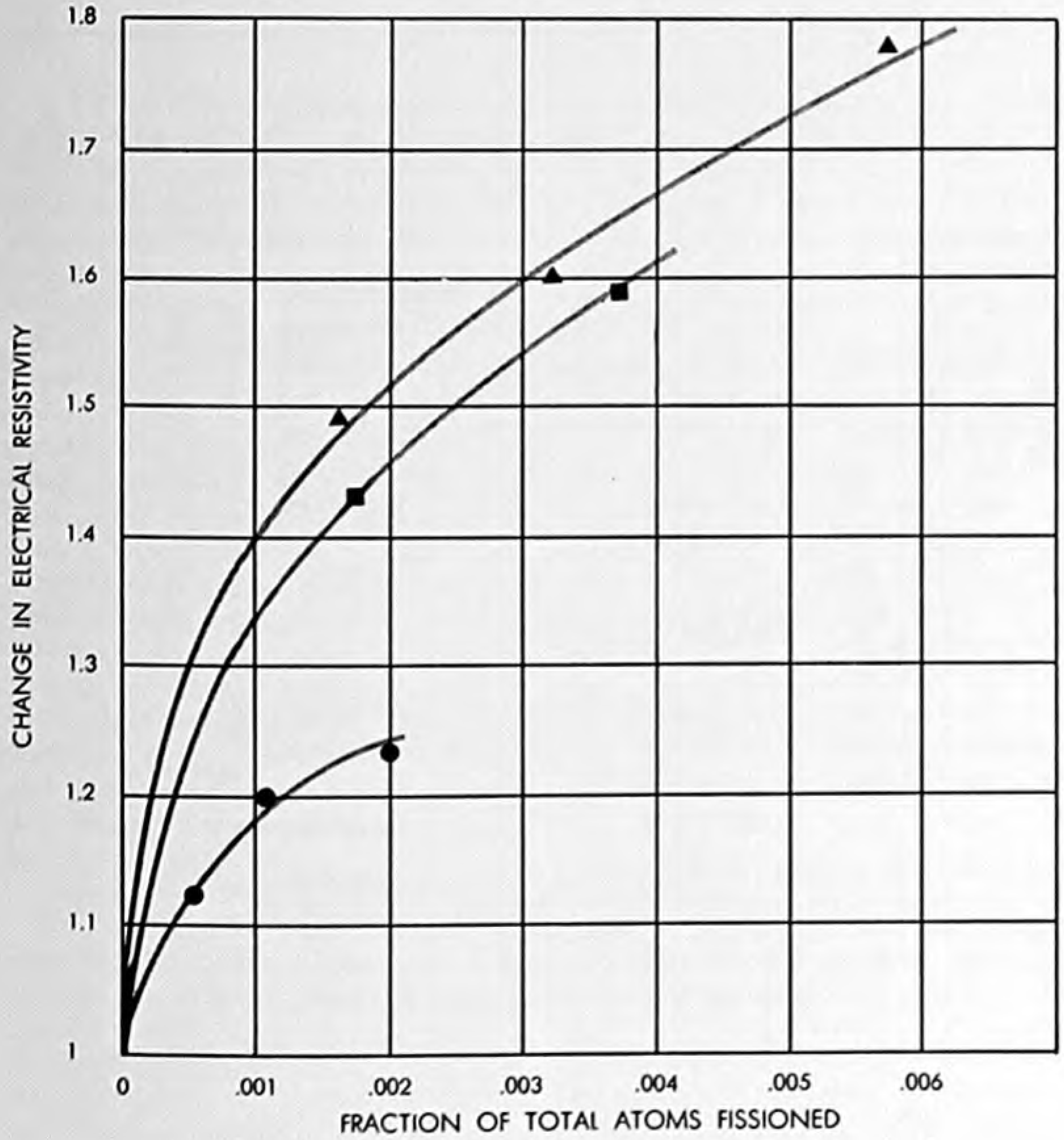


URANIUM LATTICE, also shown in two dimensions, suffers the same kind of damage as the carbon lattice, with an added complication. Sometimes the neutron is absorbed by the uranium nucleus, causing fission. The fission products may do more damage than neutrons.



DEUTERON BOMBARDMENT of copper (*broken curve*), silver (*solid curve*) and gold (*dotted curve*) increases the electrical resistance of these metals. The horizontal scale gives the number of particles per second in each 10^{-15} square centimeters of cross section of the beam. The vertical scale shows the increase in resistance, measured in ten millionths of an ohm, of a one-centimeter cube. Measurements were made at 10 degrees above absolute zero.





CHANGE IN PROPERTIES of uranium-aluminum alloys caused by fission of some of the uranium atoms was measured by D. S. Billington of Oak Ridge National Laboratory. The upper graph shows the increase in resistance to the flow of heat; the lower graph, the increase in resistance to electricity. Squares, circles and triangles represent observations on alloys containing respectively 5.7, 15 and 17.2 per cent uranium by weight. The vertical scales give the ratio of final to initial value. Thus the upper graph shows that when two thousandths of the atoms have split, the thermal resistivity is 1.2 times its normal value.

possible benefits—produced by radiation? As we have seen, the microscopic result of irradiation is the formation of lattice defects. How do these defects alter the properties of the material? There are four important kinds of predictable changes.

First, we know that properties such as conduction of electricity and heat depend on a regular and undistorted lattice. We are not surprised to find, then, that the conductivity of materials for both electricity and heat falls sharply with increasing irradiation [*see charts at the left*]. Losses of conductivity up to 30-fold have been measured. Fortunately reactors do not rely too heavily on the heat conductivity of the moderator, and that of metals is less severely affected by radiation. Hence the decreases in conductivity do not cause real concern from the point of view of reactor operation. We must quickly add, however, that these changes do affect the instruments stuck into the reactor and must be taken into account in problems concerned with instrumentation.

The second type of radiation damage is represented by a loss of ductility. The lattice defects have the effect of blocking the glide planes of the crystals. Thus the materials behave as if work-hardened, and in fact may become brittle. This damage affects the handling of uranium fuel elements and is a major cause for concern. The changes in ductility can be spectacular. The effect was demonstrated in a U. S. atomic energy display at Geneva. Every few seconds a light ball was thrown alternately at two copper cylinders, which looked identical but differed in the fact that one had been exposed to the neutrons of the Oak Ridge reactor. The normal cylinder, when hit by the light ball, gave no sound. But the irradiated one sang like a tuning fork. We understand that no amount of normal cold-working could endow copper with as much rigidity as this irradiated specimen possessed.

These first two types of effects—on conductivity and ductility—are the most striking but not necessarily the most harmful changes caused by irradiation. From the point of view of reactor operation there are two others which have caused more anxiety.

One is a swelling of the material. The displacement of atoms to irregular positions in the lattice expands the crystals. Hence the volume of a block of material increases as the dosage of radiation increases. When the Materials Testing Reactor of the AEC in Idaho went into operation with its new beryllium oxide moderator, the moderator expanded

about 1 per cent the first day. Fortunately this expansion did not proceed linearly with time: after 10 days it was much less than 10 per cent. Nevertheless, it can be very disconcerting to have to use as structural elements materials which change their dimensions after they are installed.

The other disquieting effect of radiation is an unstable energy situation. The interstitial atoms represent a considerable amount of stored energy. When they move back into vacancies in the lattice, they release this energy. The amount of sword-of-Damocles energy stored in this way can reach values up to hundreds of calories per mole (one gram multiplied by the molecular weight of the material). Obviously a sudden release of it could lead to unpleasant complications. On the other hand, this property also has constructive possibilities: some have suggested using irradiated graphite as a kind of storage battery.

We call the various effects mentioned "damage" because they change critical properties of materials that have been placed in reactors to perform definite functions based in part on these properties. The changes in properties are regarded as harmful not because they would not be useful under certain circumstances, but because they impair the behavior for which the material was selected. To minimize the effects of these changes in a reactor, it has been suggested that materials might be deliberately irradiated before they go into the reactor. This stratagem might yield materials with desired properties and stability against further irradiation.

Indeed, we can expect that irradiated materials will be put to more and more uses as understanding of their properties and potentialities grows. Graphite storage batteries and the superhardening of copper are only a beginning of the list of possibilities. We have scarcely scratched the surface of knowledge of the radiation-induced properties of materials.

Speaking as individuals who have been interested in radiation effects on solids since the conception of the first large reactors, we find it gratifying that a phenomenon which originated as a pure nuisance promises to provide us with useful information about the solid state in general and about many of the materials we use every day.

The Authors

FREDERICK SEITZ and EUGENE P. WIGNER are among the physicists who have pioneered the field of atomic energy. Seitz was born in San Francisco and got his doctorate at Princeton. He did work for the Office of Scientific Research and Development and for the Atomic Energy Commission. Since 1949 he has been professor of physics at the University of Illinois. Wigner was born in Budapest and educated in Germany. In 1930 he joined the faculty of Princeton University, where he is now Thomas D. Jones Professor of Mathematical Physics. At the Metallurgical Laboratory in Chicago he helped to design and build the first atomic pile underneath the stands at Stagg Field. Just after the war Wigner was director of research and development at the Clinton Laboratories in Oak Ridge.

Bibliography

THE MODERN THEORY OF SOLIDS. Frederick Seitz. McGraw-Hill Book Company, 1940.

SCIENTIFIC AMERICAN October 1958, Vol. 199, No. 4, pp. 28-35 OFFPRINT 246

THE STELLARATOR

by Lyman Spitzer, Jr.

Described at the Geneva conference in September 1958, this experimental device for fusion power research employs a twisted magnetic field to obtain an ionized gas and electric fields to heat it.

At the first International Conference on the Peaceful Uses of Atomic Energy in Geneva three years ago, representatives of 72 nations gathered to compare their efforts to extract economical power from the energy liberated by nuclear fission. As the Conference opened, however, fission reactions momentarily lost the center of the stage to an even more exciting prospect. The president of the Conference, Homi J. Bhabha, predicted that within 20 years man would be deriving useful power from the thermonuclear reactions. He revealed that Indian workers were already engaged in active development of this possibility. Since these reactions, involving the fusion of nuclei of the heavy isotopes of hydrogen, may look to the oceans as a source of fuel, such a technological breakthrough would guarantee mankind a practically inexhaustible supply of energy.

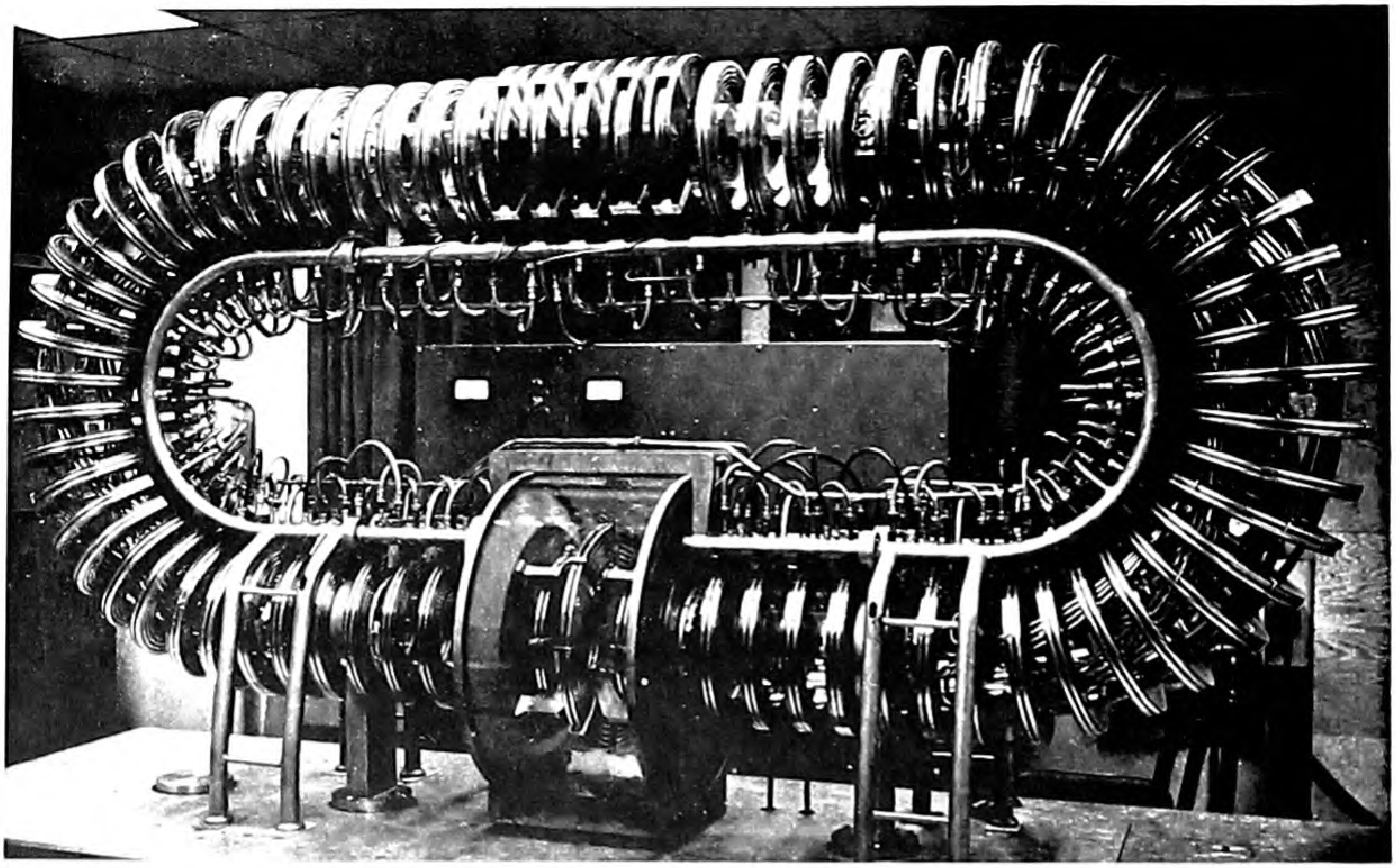
Last month, at the second Geneva conference on atomic power, the fusion reactions were at the center of the stage to stay. The knotty and often profound questions encountered in the research engaged a substantial portion of the formal and informal discussions. The elaborate experimental gear exhibited by the United Kingdom, the U. S. and the U.S.S.R. testified to the huge scale of the programs the major nuclear powers have been conducting, until recently in secret. Amid the hopeful accounts of progress, no one could yet claim conclusive evidence for the attainment of a controlled thermonuclear reaction. On the other hand, no one reported the discovery of any insurmountable obstacle.

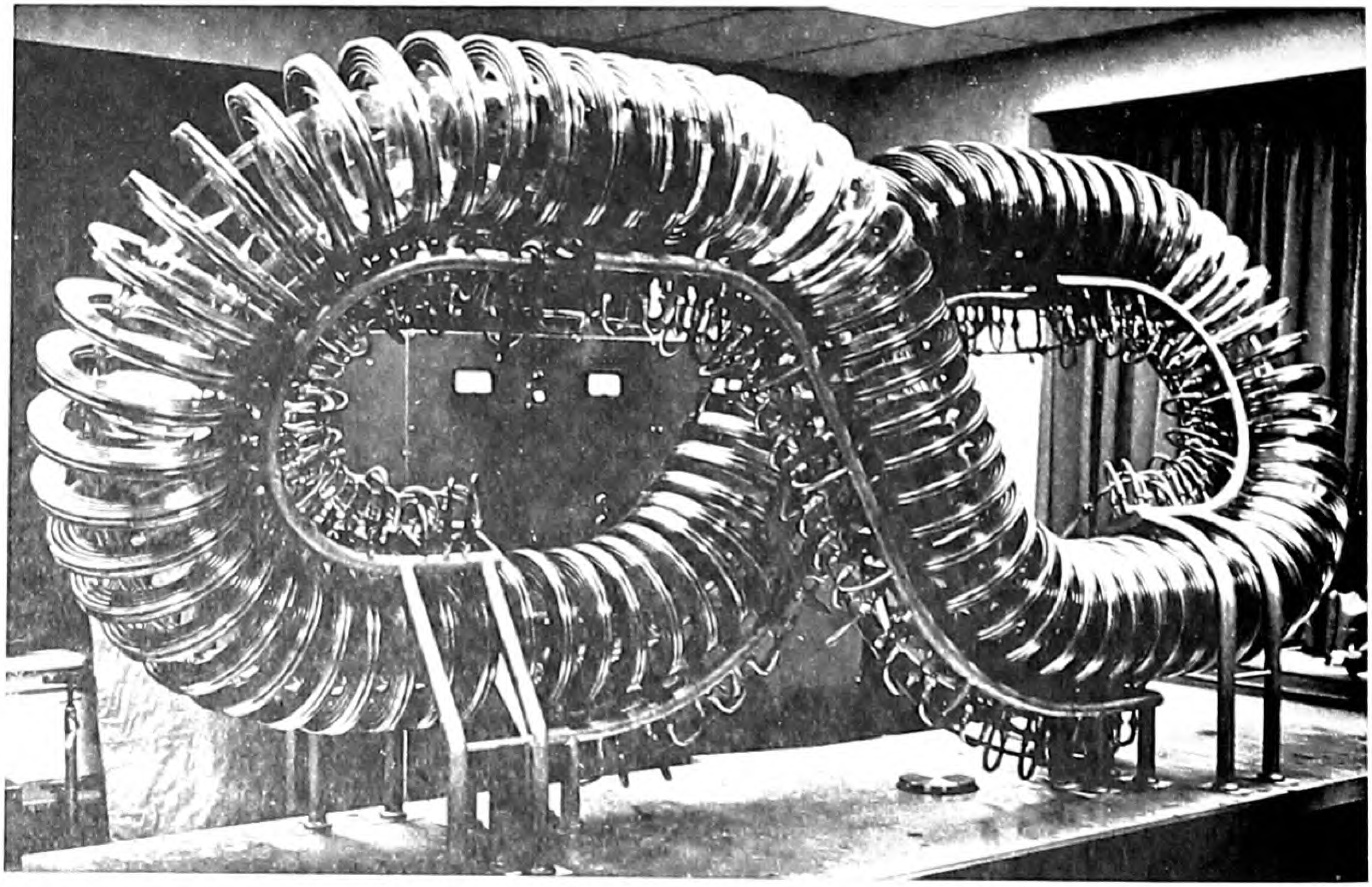
Among the several approaches of the U. S. program publicly disclosed in detail for the first time at Geneva is the "stellarator," the subject of this article. It is the work of Project Matterhorn at Princeton University and embodies some of the things we have learned from theoretical and experimental investigations conducted since 1951.

For us, as for others in the field, the work is defined by the dual objective of igniting and containing a thermonuclear reaction. When two nuclei of heavy hydrogen collide at high energies, they interact, forming a new nucleus and liberating either a proton or a neutron of high energy. To obtain a useful power yield from this reaction, a gas of deuterium (hydrogen of mass 2) or of mixed deuterium and tritium (hydrogen of mass 3) must be brought to an enormously high temperature—about 100 million degrees absolute (degrees centigrade above absolute zero). The problem then is to contain or confine the gas within a limited region away from contact with any solid material. If the ultrahot hydrogen strikes anything which is solid and, of necessity, relatively cool, the temperature of the gas will drop sharply, and the solid material will tend to evaporate.

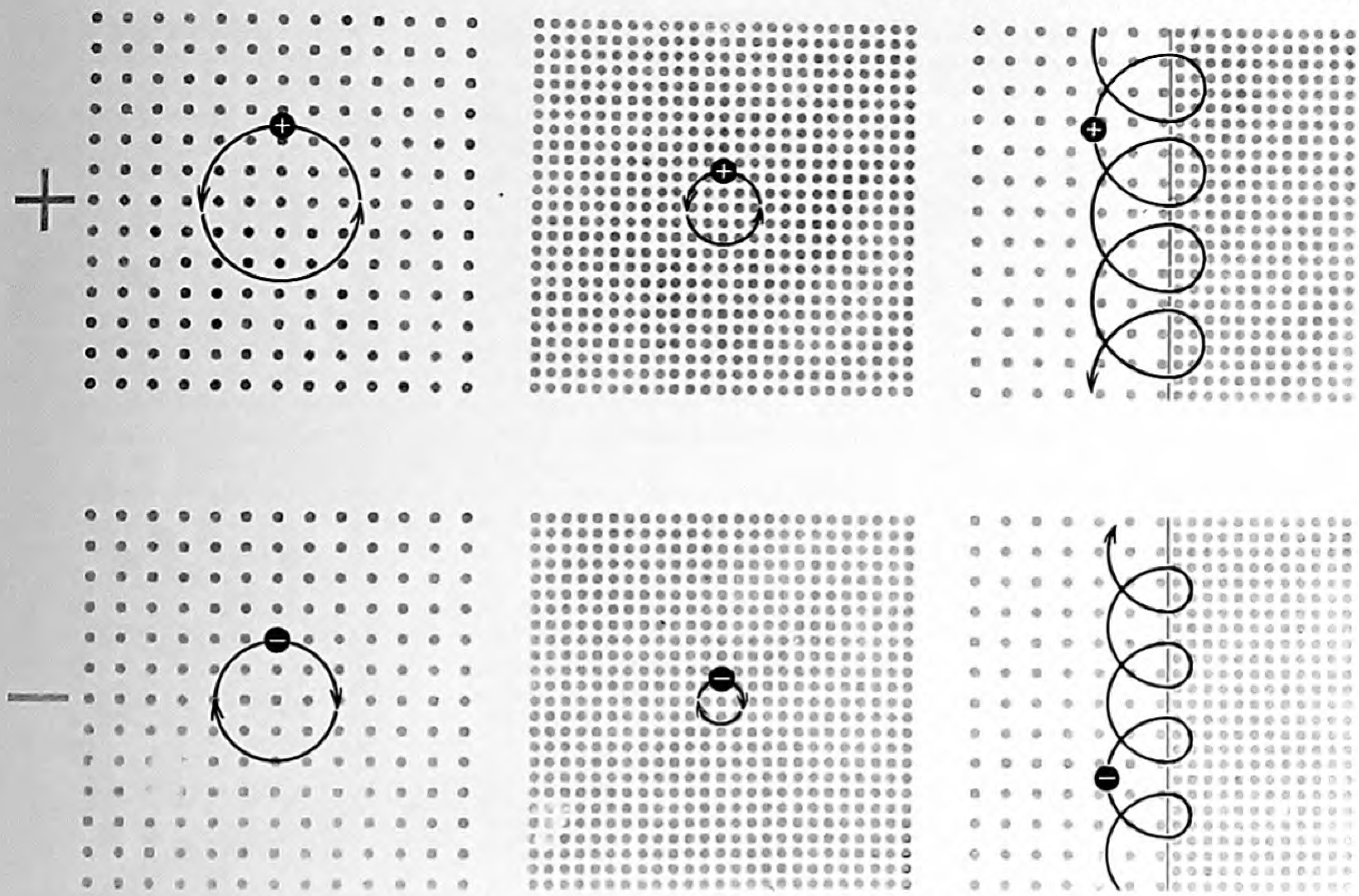
Magnetic forces seem to offer the only way to contain a thermonuclear reaction. One effort in the search for a "magnetic bottle" has been devoted to the so-called "pinch effect." Here the flow of a heavy electrical current through the hot gas generates a strong magnetic field which at once contains the gas and brings it up to high temperature by compressing, that is, pinching, it [see "Fusion Power," by Richard F. Post; SCIENTIFIC AMERICAN Offprint 236]. The Matterhorn project is trying a somewhat different tack. In the stellarator we are undertaking to contain the gas in a magnetic field produced independently of the electric current that heats the gas up through the first million degrees. We hope to reach ultrahigh temperatures via an effect called "magnetic pumping," induced by a second externally generated, rapidly pulsating magnetic field. Whether the stellarator will achieve a controlled thermonuclear reaction we do not yet know. At each step in its development we have had to solve unexpected problems, and each time have learned something new about the weird and wonderful properties of a hot gas in a magnetic field. Many more such problems await us.

The stellarator produces its containing magnetic field by the perfectly straightforward method of passing a strong electric current through a solenoidal coil [*see diagram at top left on page 384*]. In the familiar example of a straight, cylindrical coil, such as that employed in simple relays, the magnetic field parallels the coil axis and has the same strength everywhere both along its





STELLARATOR TUBES, photographed at Project Matterhorn at Princeton University, have twisted magnetic fields designed to prevent the contained gas from striking the vessel walls. In the arrangement at top the twisting is produced by means of the set of helical coils which can be seen between the sections of the main field coil. Figure-eight shape at bottom also produces twist.



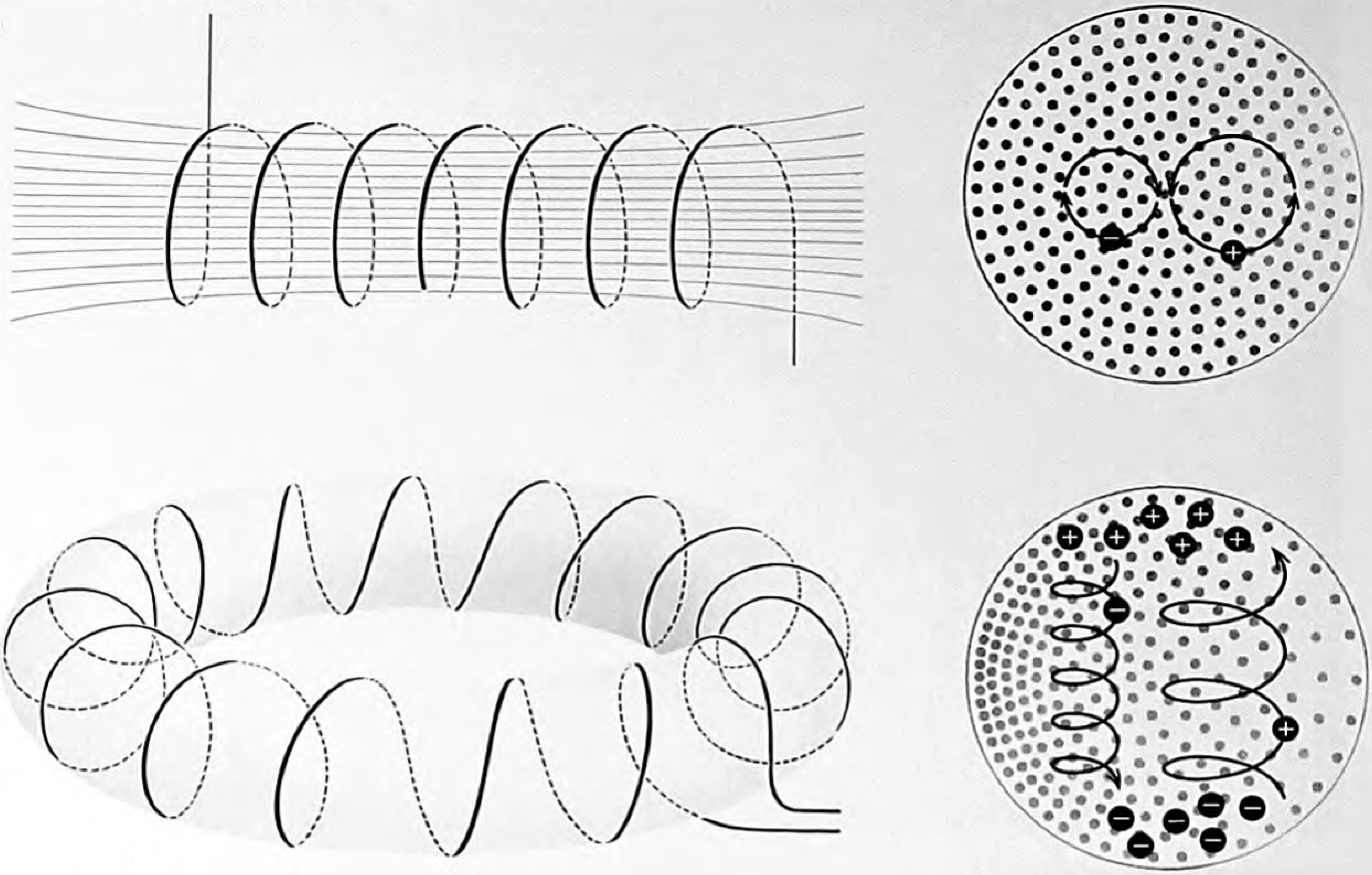
PATHS OF CHARGED PARTICLES in magnetic fields are shown schematically. Small dots represent cross sections of lines of force. In a weak, uniform field (*left*) a positive particle such as a deuteron ion (*top*) describes a circle in one direction; a negative particle such as an electron (*bottom*), a circle in the opposite direction. (Diameter of deuteron's path is actually 60 times as large as electron's.) In strong field (*second from left*) the circles are smaller. In an imaginary field changing abruptly from weak to strong (*right*), paths are spirals in opposite directions, made up of semicircles from the large and small circular paths.

length and through its cross section. The "lines of magnetic force," the imaginary lines following the direction of the field, are straight well inside the coil. If we place a hot deuterium gas in a tube inside such a field, it will respond to the magnetic force because its atoms are ionized, *i.e.*, stripped of their electrons. The electrically charged particles (positively charged nuclei and free electrons) moving across the magnetic field experience a deflecting force, and proceed to gyrate in circles about the lines of magnetic force. Since the motion of the particles along the lines of force is unaffected by the field, the path of each particle is a helix. Thus contained in the magnetic field, the gas will not touch the walls of the tube. We have here, in fact, an ideal bottle for a fusion reactor, were it not for the ends of the coil, where the lines of force must emerge and where the gas in consequence must come in contact with solid matter.

The simplest way to get rid of the ends of a long coil is to bend the coil in a circle, joining the two ends together to form a closed but endless tube. In such a toroidal, or doughnut-shaped, coil [*see diagram at bottom left on following page*] lines of force become circles. Unfortunately this circular coil has a fatal defect which keeps it from qualifying as a satisfactory magnetic bottle. As a result of curvature, the strength of the magnetic field is greater near the inside wall of the tube than it is near the outside. This inhomogeneity of the field alters the helical path of the charged particles. Near the inside wall the relatively stronger field curves the path of a particle more sharply than near the outside. The result is that the charged particles drift across the field, the positively charged nuclei collecting at the top of the tube and the electrons at the bottom [*see diagram at bottom right on page 384*].

This drift is bad enough by itself, but its indirect effect is catastrophic. The resultant separation of electric charges produces a large electric field, which disrupts the particle paths completely, throwing the entire gas into the wall. If the electric charges could only leak back across the lines of force, this disaster could be averted, but in a fully ionized gas this is not possible. It is a remarkable fact that a steady electric field imposed across a magnetic field produces no current at all in a fully ionized gas, but drives the gas particles in a direction at right angles to both the electric and magnetic fields. A steady electric field imposed along the magnetic field will easily produce an electric current parallel to the lines of force. But this does not help matters in the case of the simple torus, since the lines of force do not connect the regions of opposite electric charge.

It happens that there is a simple remedy for this catastrophic drift of charged particles across the toroidal magnetic field. By one means or another we can twist the magnetic field around



MAGNETIC CONTAINMENT of charged particles can be accomplished with a uniform field like that inside a long, straight tube wound with a helical coil carrying electric current (*top left*). Particles will travel through the tube in helical paths made up of circular motions around lines of force (*top right*) and motion along the lines. If the ends are closed by bending the tube into a doughnut (*bottom left*), the field is no longer uniform, being stronger on the inner part of the doughnut (*bottom right*). The result is that positive and negative particles separate, spiraling to top and bottom as in the right-hand diagrams on the preceding page.

its circular axis (called the "magnetic axis"), giving the lines of force a helical form like the strands of a rope. In this configuration a single line of magnetic force, if followed indefinitely on successive turns around the torus and successive twists around the magnetic axis, will trace out an entire surface, called a "magnetic surface." Lines of force near the magnetic axis will produce magnetic surfaces which are nested inside the surfaces produced by the lines of force farther out. This family of nested magnetic surfaces constitutes the magnetic bottle proposed for confining an ultrahot gas in a stellarator.

In this twisted toroidal field the effect of particle drift is much reduced. Most particles are moving rapidly along lines of force, and so are rotating about the magnetic axis as they follow the twist in the lines. When the upward-drifting particles are above the magnetic axis, they of course tend to move farther from the axis; when they are below, however, the same upward drift compensates for this by moving them back toward the axis. As a result their average distance from the magnetic axis does not change. Oppositely charged particles still show some tendency to drift apart, with an accompanying separation of charges. But now the charges can leak back along the lines of force, since a single line, if followed for a distance, leads from the top of the tube to the bottom. Any differences in electric charge along a line of force are thus eliminated, and a steady confinement of the ionized gas now becomes possible.

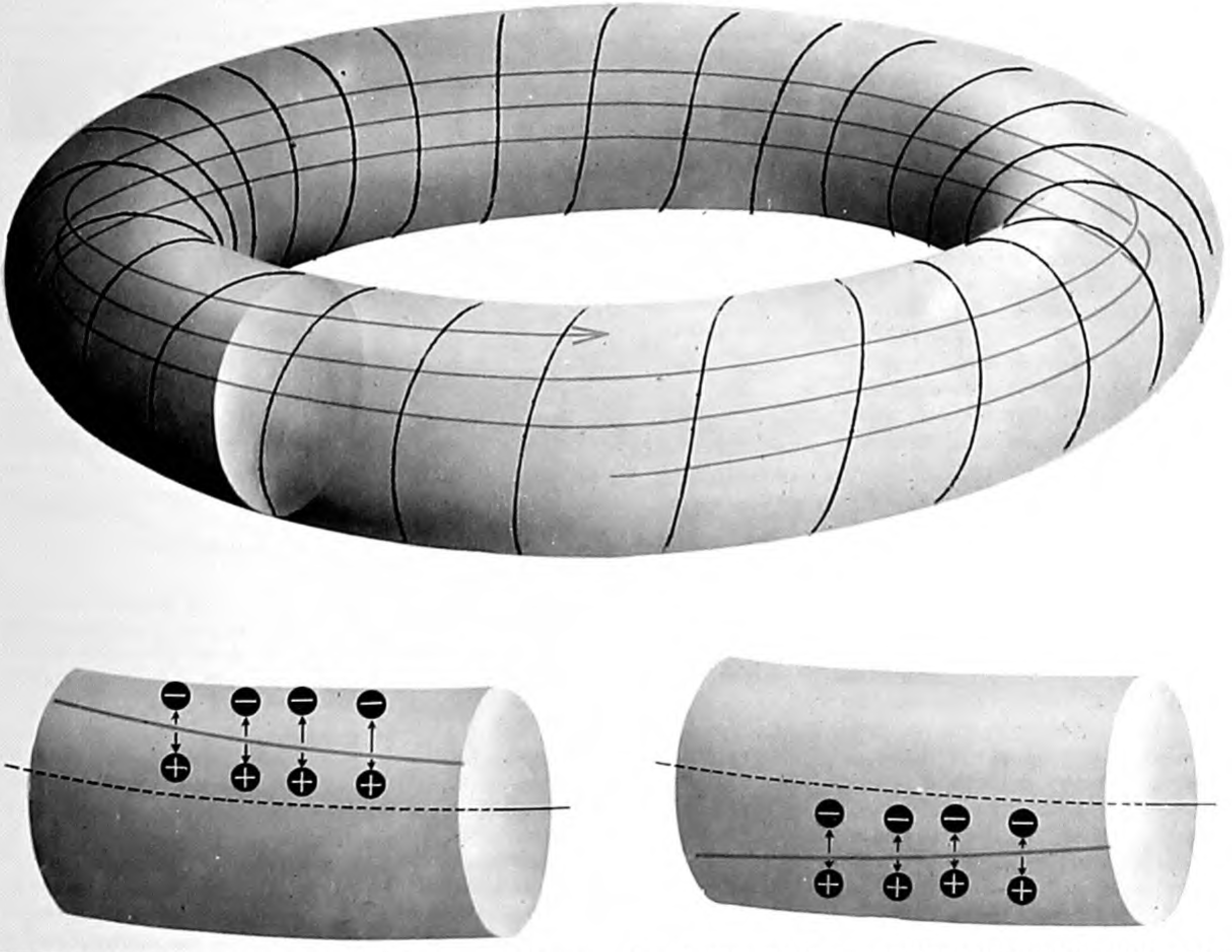
The necessary twist can be imposed on the toroidal field in a number of ways. Passing an electric current along the lines of magnetic force in a torus will do it, but such a current cannot be maintained in a steady state, and would require pulsing every few seconds. Since it is desirable to operate a fusion reactor continuously, without pulsing, some other method of twisting the lines of force is preferable. It turns out that merely bending the simple torus into a figure eight will produce an appreciable twist in the lines of force; most of the experimental work at Project Matterhorn has been carried out with figure-eight-shaped magnetic fields. We are now developing a more promising method in which the toroidal field is twisted by interaction with an additional transverse magnetic field. This transverse field is generated by a set of helical windings in which the current flows in opposite directions in adjacent groups of wires. In the twisting of the field thereby induced, the angle of twist increases as the square of the distance from the magnetic axis; this difference in twist angle between the outer and inner magnetic surfaces helps to stabilize the gas.

At this point we may consider the magnetic bottle of the stellarator more or less ready to contain the hot gas, and we may turn to the question of heating the gas to high temperatures. The stellarator accomplishes the preliminary

phase of heating simply enough by inducing a strong electrical current in the gas. With an iron transformer core threading the toroidal tube and a primary winding wrapped around the iron, we have a conventional transformer, the gas providing the secondary circuit. When a voltage is applied to the primary winding, an electric current flows through the gas along the magnetic surfaces. The power dissipated by the resistance of the gas goes into ionizing and heating the gas, as well as into producing some ultraviolet and visible radiation. This is called "ohmic heating," because it is the ohmic resistance of the gas that generates the heat on passage of the current. Unlike the pinch-effect current, the ohmic heating current produces no contraction or compression of the ionized gas; the strong magnetic field of the stellarator holds the gas firmly in place and constant in volume.

Experience with some half-dozen small "Model B" stellarators, with tubes of two to four inches in diameter and 10 to 20 feet in axial length, has established extensive data on ohmic heating. These devices utilized magnetic confining fields of 20,000 to 40,000 gauss. Since the power required at the peak of the field is in the neighborhood of 50,000 kilowatts, the power bill has restricted operations to pulses lasting about .02 second. Within these intervals we have been able to develop ohmic heating discharges of a few thousandths of a second duration. Temperatures of 100,000 to a million degrees observed in these experiments sustain the expectation that ohmic heating is an effective way to ionize the gas in a stellarator and heat it to high temperatures.

In the early days of the program we hoped that the gas would remain steady and motionless during ohmic heating, and that the heating would thus proceed smoothly to its conclusion. Experiments have shown us that this hope was about as well founded as the hope that water flowing rapidly through a large pipe would flow smoothly in straight lines. In truth, of course, water under these conditions becomes turbulent, with eddies moving back and forth across the stream. In the same way the ionized gas develops violent activity during the ohmic heating process. Electric and magnetic forces introduce complications which make the activity quite different from ordinary turbulence. Because of the long-range electrical forces between electrons and the bare nuclei, particles considerable distances from each other can cooperate in large-scale activity. This "cooperative activity" in a heated gas produces many effects. Not all are disruptive; for example, the production of radio-noise bursts similar to those observed from the sun does not appear to affect the performance of a ther-

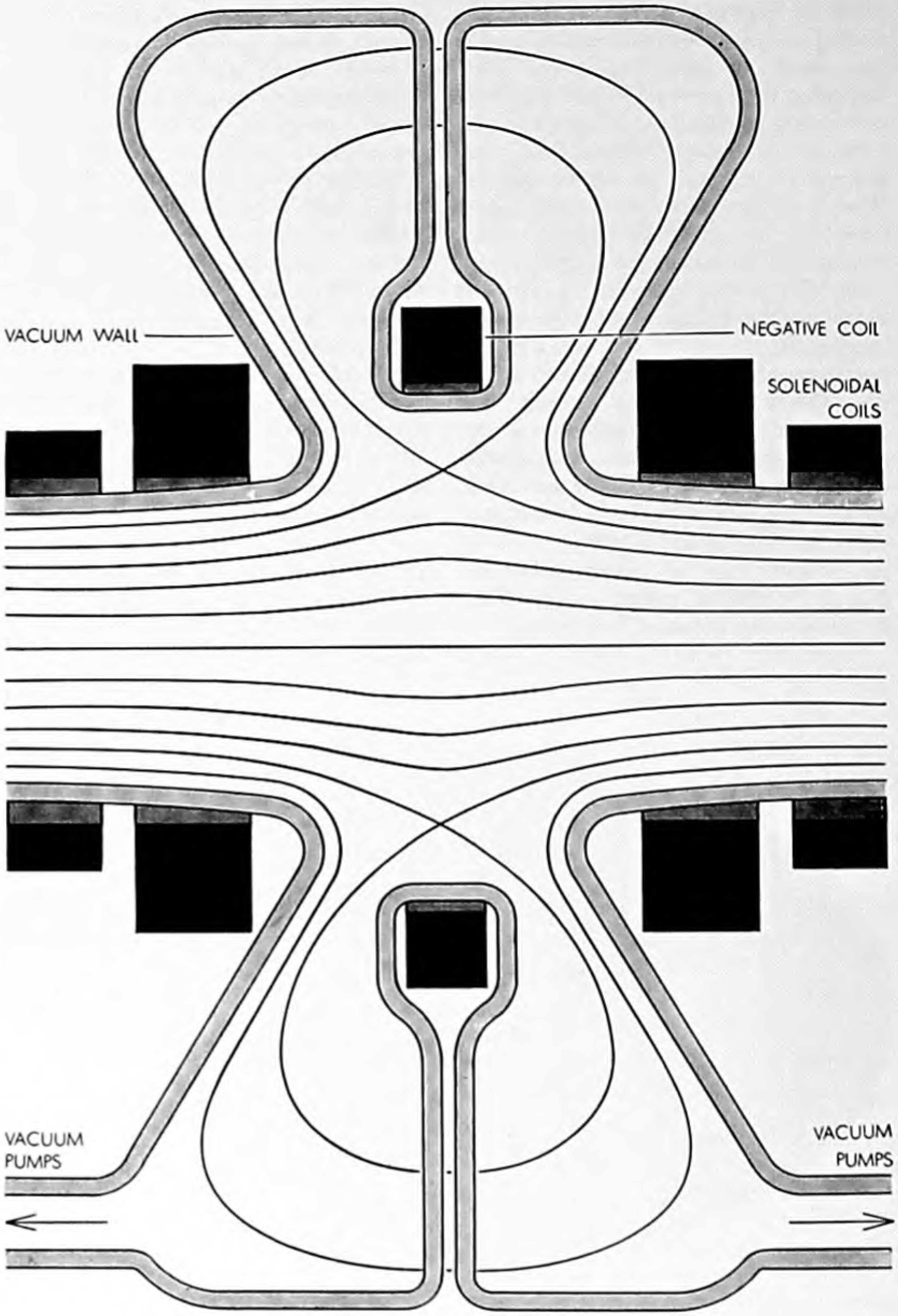


TWISTED MAGNETIC FIELD counteracts tendency of charges to separate. Top diagram shows a part of one line of force, which, if drawn for its whole length, generates a ring-shaped surface. Each line is above axis in some sections of tube (*bottom left*) and below it in others (*right*). Thus both plus and minus charges move away from center line at some points and toward it at others.

monuclear reactor. Other effects may have very serious consequences. Some types of cooperative activity move the individual particles across the lines of force, impairing confinement by the magnetic field and dashing some of the hot gas into the wall. In activity of the so-called "hydromagnetic" (or "magneto-hydrodynamic") type the lines of magnetic force themselves become violently distorted, also dashing ionized gas into the wall. The tendency to such hydromagnetic disturbance should be much reduced in stellarators in which the magnetic field is twisted by transverse helical fields. Theory indicates that the lines of force cannot move about so readily when the angle of twist increases sharply with distance from the magnetic axis.

"Runaway" electrons seem to be responsible for another troublesome variety of cooperative activity. One characteristic of an ionized gas is that the probability of a collision between an electron and another particle falls off rapidly as the particle velocity increases. Thus an electron accelerated to a sufficiently high velocity will engage in practically no collisions and will continue to be accelerated by the field. Such electrons are said to have run away; they can no longer be confined and they strike the wall, producing intense X-rays. When a large number of the electrons runs away, the gas develops strong fleeting electrical fields and the current drops more or less abruptly.

Incidentally, observations on runaway electrons provide an excellent demonstration of the effectiveness of the stellarator magnetic bottle. Under some conditions energetic electrons will travel around the stellarator long after the ohmic heating pulse is over, finally striking the wall only when the magnetic field falls off. During this time they travel several thousand miles, going around the stellarator tube about a million times.

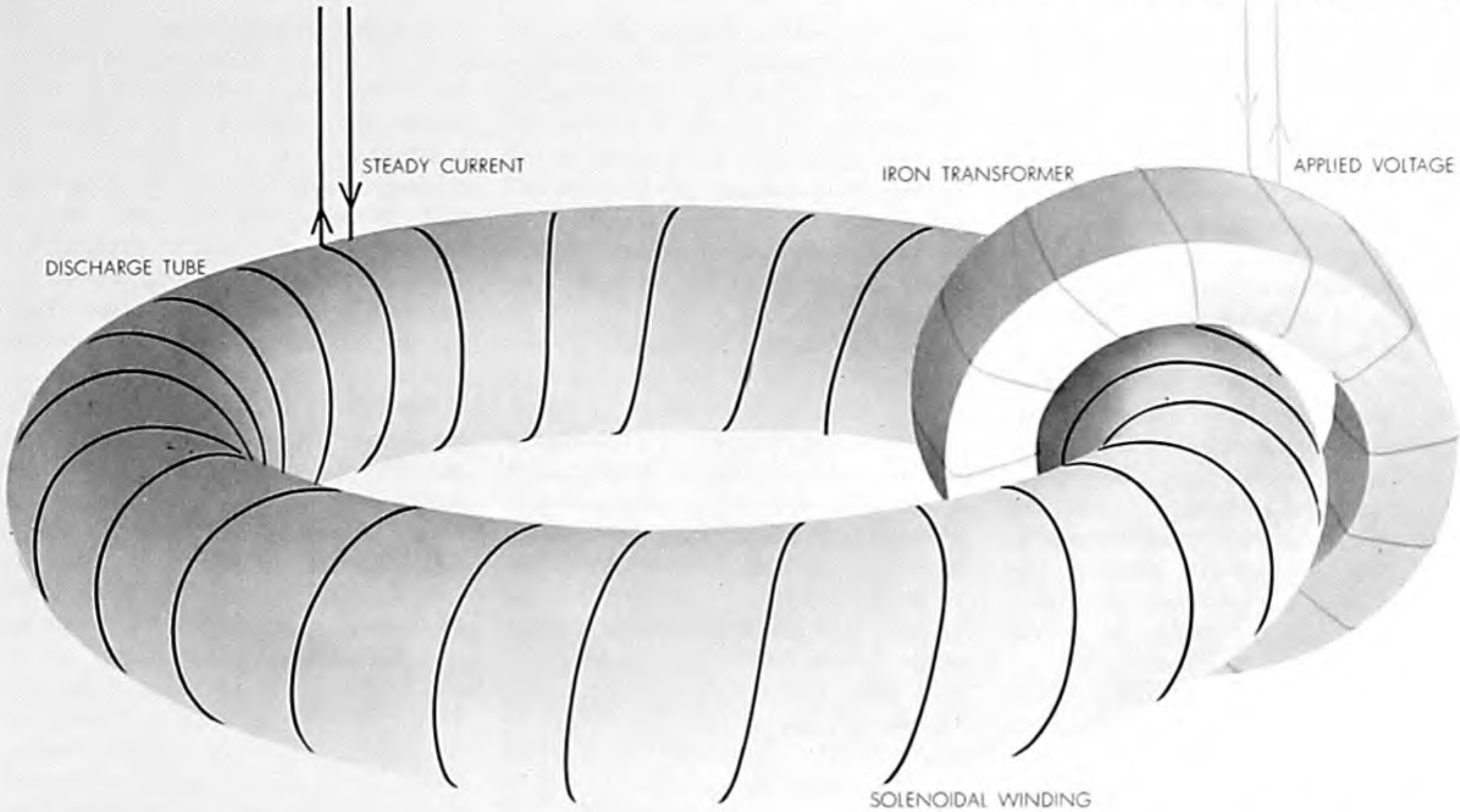


DIVERTOR helps prevent gas particles from striking walls of tube. Magnetic lines near the walls are bent out into side chamber surrounding tube at one point. Particles traveling along these lines are swept into the chamber and pumped out of the system. This diagram is a cross section of the divertor; the black blocks are coils that bend the lines.

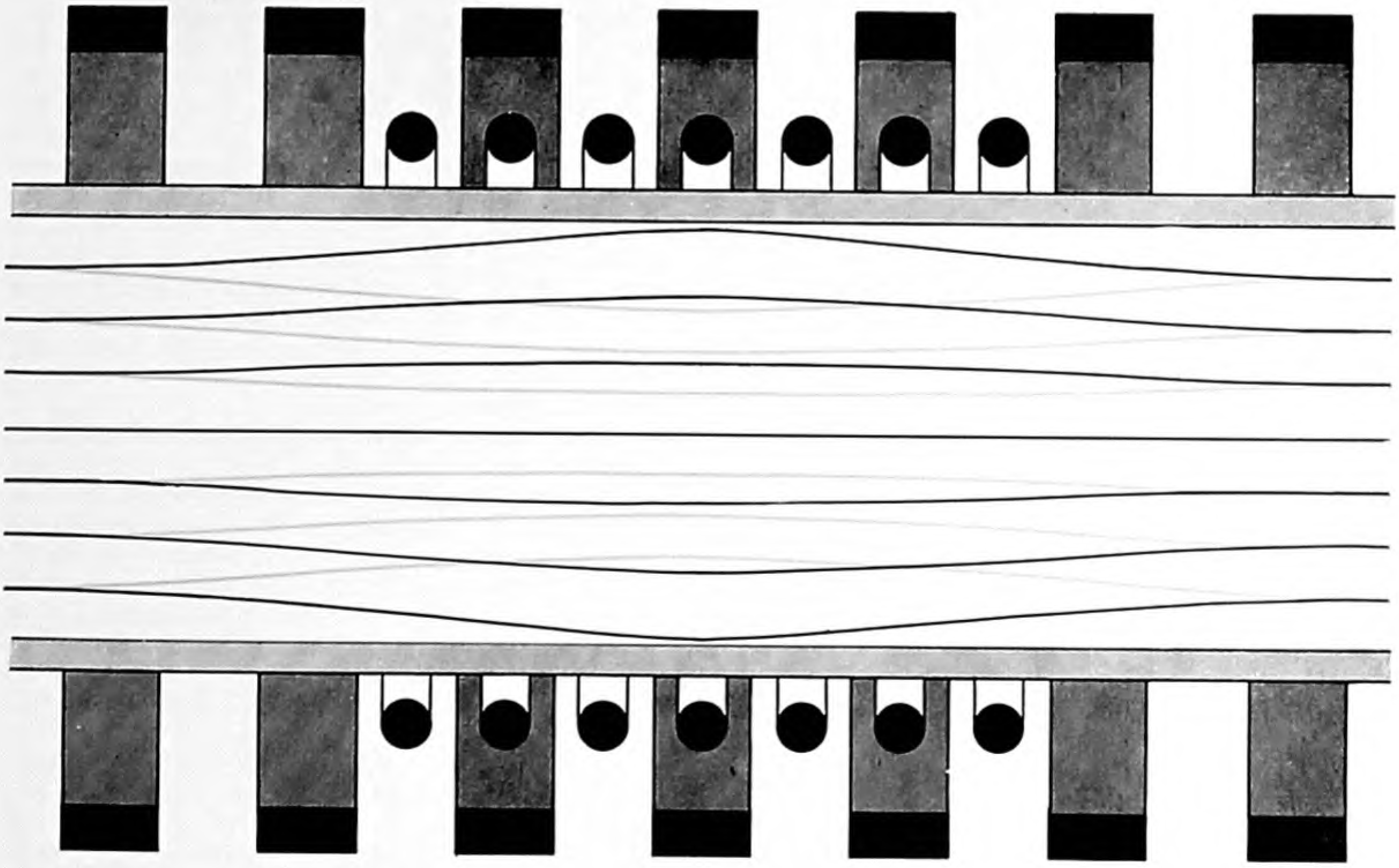
The most serious type of cooperative activity is not really understood at all. On occasion, with the gas showing no strong evidence of unsteadiness, the ions and electrons gradually disappear from the discharge, presumably striking the wall. This disappearance of the gas, called "pump-out," represents a serious leak in the magnetic bottle. When ohmic heating terminates, the leak seems to stop, as is evidenced by the long persistence of runaway electrons. A program of investigation to determine the cause of this leak is now under way.

A leak might not be serious if a particle striking the wall simply buried itself there. At least the gas remaining in the bottle would stay hot. Unfortunately an ion striking the wall tends to knock off atoms of various elements. These find their way into the heated gas and cool it, partly because they are themselves so cool to begin with and partly because they radiate energy at such a rapid rate. A controlled thermonuclear reactor is possible in principle only because hydrogen at high temperatures radiates so little energy, and can therefore be kept very hot. Atoms of the heavier elements do not possess this convenient property, and because of their higher nuclear charge radiate energy at a much greater rate. Thus any appreciable amount of oxygen or iron in the discharge makes it very difficult, if not impossible, to maintain thermonuclear temperatures.

Two methods give promise of reduc-



TRANSFORMER ACTION heats gas in stellarator to a temperature of about one million degrees. Current through primary coil around iron core at right induces a secondary current through the gas in the tube. As in an ordinary metallic conductor, the resistance to the flow of current produces heat. As gas gets hotter its resistance drops until further heating becomes impossible.



MAGNETIC PUMPING will be used to heat gas to temperature necessary for fusion reaction. Pumping arrangement, shown in cross section, consists of an auxiliary coil (*black circles*) carrying an oscillating current. This current alternately expands and compresses the lines of force set up by the main coil (*black blocks*). Diagram shows compressed field in color and expanded in black.

ing the influx of impurities from the wall. The chief impurities observed are carbon, oxygen and nitrogen, which are adsorbed on the steel walls of the stellarator tube, and which stream from the walls during the discharge as a result of wall bombardment. To reduce the amount of these gases present, the stellarator tubes are now baked at 450 degrees C. for some 12 hours while the vacuum pumps are operating. This has reduced the influx of impurities during ohmic heating to about a hundredth of its value in unbaked tubes. Baking the vessel has the auxiliary advantage of reducing the initial vacuum (before the discharge gas is introduced) down to 4×10^{-13} atmospheres, about a ten-thousandth of that obtained with more usual techniques. To reduce further the impurities produced during operation, the next generation of stellarators will mostly be equipped with what we call divertors. At the divertor the outer shell of magnetic lines of force is bent away from the axis of the field [*see illustration on page 386*]. Any charged particles moving along these lines of force, and hence tending to diffuse toward the wall, are drawn into the separate chamber of the divertor before they strike the wall. Impurities produced in the divertor are scavenged by the vacuum pumps. Use of a divertor on an unbaked system has reduced the impurity level to about a fifth of its previous value. In a larger, baked machine use of the divertor may bring the impurities down to a negligible value.

Quite apart from problems of cooperative activity, experience shows that ohmic heating becomes less and less effective as the gas temperature rises above a million degrees. The resistance of the ionized gas decreases sharply as the temperature rises. Since the power dissipated by an electric current depends upon the resistance, the power dissipation drops proportionately. Some other method must be found for heating the gas above a million degrees.

At Project Matterhorn we propose to heat the gas up to ignition temperature by "magnetic pumping," a rapid pulsation of the containing magnetic field along a short section of the stellarator. This pulsation is produced by an oscillating current in a special coil surrounding the heating section, inside the coils which produce the steady magnetic field [*see illustration at bottom of the preceding page*]. As the field strength increases during the pulsation, the lines of force move closer together, and the charged particles, following the lines of force, also move together. The gas is thereby compressed; with compression, it is heated. The details of this process are complicated, but the net effect on the gas temperature is the same as the heating of an ordinary gas by compression. Hence we may look on the moving lines of magnetic force as a piston which compresses and expands the ionized gas.

Although the gas is heated by compression, it is also cooled by expansion. It is not immediately obvious, therefore, why there should be any net heating. In fact, in order to achieve heating it is necessary to tune the frequency of pulsation close to one of the natural periods of the gas—the time between collisions, or the time it takes a charged particle to pass through the heating section, or the time it takes a particle to gyrate in a circle under the influence of the confining field. Thanks to a type of resonance, the amount of heat fed into the gas during compression will then considerably exceed the heat taken out during expansion. Detailed theoretical studies, still unchecked by experiment, indicate that magnetic pumping should be capable of heating a gas to thermonuclear temperatures. The rate of heating by this technique, however, is not very rapid, and so far the loss of energy by pump-out and the radiation of energy by impurities have prevented the achievement of ultrahigh temperatures.

A much larger experimental device, the "Model C" research stellarator now under construction, should aid greatly in the investigation of these problems. This device will have a vacuum tube in the shape of a racetrack; the tube will be some eight inches in diameter and 40 feet in axial length, and will utilize a confining field of 50,000 gauss, twisted by means of helical windings. The tube diameter is some four times larger than in most of the present "B" models, and the correspondingly decreased ratio of surface to volume should reduce the effect of pump-out and of the impurities streaming into the discharge. Finally, powerful magnetic pumping apparatus will be available for attempts to obtain ultrahigh temperatures. The engineering facilities required for this installation are substantial, with a peak direct-current power level of 150,000 kilowatts, obtained with motor-generator sets, and a radio-frequency power level of about the same order.

If continuing research does lead to a power-producing reactor, what would such a device look like? The main answer to this question is that a stellarator must be a very large unit to produce more power than is consumed in maintaining the magnetic field needed for confinement. As a stellarator is increased in size, keeping the magnetic-field strength and all other properties constant, the power required for the magnetic field increases only as the linear dimension. On the other hand, the thermonuclear power produced increases with the volume of the gas, or as the cube of the linear dimension. The break-even point, at which the power put into the magnetic field equals the thermonuclear power generated, occurs with a rather large machine—one with a tube diameter of several feet and an axial length of several hundred feet. The total power produced by such a machine is in the million-kilowatt range.

The first power-producing stellarators would be fueled with a mixture of deuterium and tritium, since tritium fuses with deuterium about a hundred times more rapidly than deuterons fuse with each other. To take the power out the reaction tube would be wrapped with a blanket through which water would be circulated in pipes. The hydrogen in the water would take up the energy of the neutrons in elastic collisions, and the water, acting as a coolant, would carry the heat out of the stellarator to external turbogenerators. To replenish the supply of tritium the blanket would be loaded with lithium; one of the isotopes of this element absorbs neutrons readily, and the ensuing nuclear disintegration yields tritium and an alpha particle (nucleus of the helium atom). Surrounding the blanket would be enormous coils in which electric currents would produce the steady magnetic field that is required to confine the ionized gas.

Once the gas is heated, such a device would operate continuously. Fresh supplies of deuterium and tritium would be injected in a fast jet, while the reaction products (mostly helium) and impurities would be withdrawn at the divertor, where very large pumps would be located. Ohmic heating and magnetic pumping equipment would be needed only for starting the stellarator after occasional shutdowns.

A full-scale stellarator would be comparable in size and power output to a large hydroelectric plant, such as Hoover Dam. Whether such an installation will ever be economically feasible, or even possible, is still uncertain. If it works, however, it will provide millions of kilowatts with a negligible fuel cost.

The Author

LYMAN SPITZER, JR., is director of Project Matterhorn, a large-scale investigation at Princeton University into the possibilities of fusion power. He is largely responsible for launching the project: in 1951, unaware that the Atomic Energy Commission was already conducting fusion research at the Los Alamos Scientific Laboratory, he approached the Commission with proposals which led to the setting up of a second research center. A native of Toledo, Ohio, Spitzer has studied at Yale, Harvard and Princeton universities and at the University of Cambridge. He has been a member of the Princeton faculty since 1947; in 1952 he was appointed Charles A. Young Professor of Astronomy, succeeding the eminent astronomer Henry Norris Russell. In addition to his duties as director of Project Matterhorn, Spitzer heads the Princeton department of astronomy and is director of the University observatory. Most of his work has been in theoretical astrophysics, though he did sonar research for the Navy during World War II. He became interested in fusion power early in 1951, after reading about the alleged Argentine success in the field. "Shortly after the Argentine story broke," he writes, "I went to Aspen, Colorado, for a ski trip. In the relaxed atmosphere there, and especially while riding the chair lift, I gave some thought to the problem of how a controlled thermonuclear reactor might be built. Many of the essential ideas of the stellarator stemmed from my thinking during that week, and during the weeks in Princeton immediately following. I don't know quite what this story proves, but perhaps it is an argument for more frequent vacations."
